DICAL ROOM
AL LIBRARY
OF MICH.

OCT 3 1932

VOL. XXIII SEPTEMBER 1932 NO. 6

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

HAROLD RUGG, *Editor*

CONTENTS

A Standardized Test of Prose Appreciation for Senior High School Pupils 401
HERBERT A. CARROLL

The Reliability of a Single Test Item 411
KARL J. HOLZINGER

A Method for Determining the Absolute Affective Value of a Series of Stimulus Situations 418
PAUL HORST

Visual Imagery and "Observation" 441
HENRY BOWERS

"Students" Method for Measuring the Significance of a Difference between Matched Groups 446
MORDECAI EZEKIEL

Standardization of the Stenquist Mechanical Assembling Test Series III 451
EDNA WILLIS MCELWEE

A Modification of the Doolittle Method: A Logarithmic Solution . . 455
R. J. WHERRY

Nomogram for Blakeman's Test for Linearity of Regression 460
HAROLD D. GRIFFIN

The Effect of Practice on the Homogeneity of a Group 462
ROBERT P. CARROLL

Discussions
Statistics Never Equivocate 465
CURT ROSENOW

"The Effect of the 6-22-44-22-6 Normal Curve System on Failures and Grade Values"—A Comment 466
WALTER CROSBY EELLS

Book Reviews 469
Publications Received 478

$5.00 per Year • Published Monthly Except June to August by

WARWICK & YORK, INC.

BALTIMORE, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

For Teachers who would guide with Intelligence rather than with Emotion

NORMAL YOUTH

And Its Everyday Problems

By DOUGLAS A. THOM, M.D.

Author of "Everyday Problems of the Everyday Child"

In his new book Dr. Douglas A. Thom discusses fully and frankly the difficulties with which parents and teachers of adolescents have to deal. He treats adolescence as a normal phase of development and lays down principles of guidance which tend to prevent these normal reactions from becoming abnormalities. With many case studies drawn from actual experiences in school, college, home and industries. $2.25

This is an Appleton Book

D. APPLETON AND COMPANY
35 West 32nd Street New York City

The Psychology of Learning
1928 Revision By WILLIAM H. PYLE
$2.20+12¢ postage
Warwick & York, Inc., *Baltimore, Maryland*

The Acid Test

What more exacting test can there be of dictionary leadership than the judgment of the educational world?

Every State that has adopted an unabridged dictionary for use in its schools has chosen exclusively

WEBSTER'S NEW INTERNATIONAL DICTIONARY

The Colleges voted overwhelmingly in favor of Webster as the standard of pronunciation, in answer to questions submitted by the Chicago Woman's Club.

Nearly 100% of all schoolbooks are based on the New International for pronunciation, spelling, compounding, and division of words.

The New International has been universally accepted by those best fitted to judge.

It is in accord with the best modern practice.

Use it in your own schools.

Write for our free booklets of interesting ways to teach the use of the dictionary.

GET THE BEST

G. & C. MERRIAM COMPANY
SPRINGFIELD, MASS.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY
ESTABLISHED 1910

Editor: HAROLD RUGG

Board of Associate Editors

J. CARLETON BELL ARTHUR I. GATES RUDOLF PINTNER E. L. THORNDIKE
FRANK N. FREEMAN V. A. C. HENMON L. M. TERMAN GOODWIN B. WATSON

The price of the Journal is $5.00 a year in the United States; $5.30 in foreign countries. Subscriptions for less than a year are charged at the rate of 70 cents for each number ordered. Back volumes are $6.00 each; back issues, 90 cents each except when three or more years old, and then $1.00 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Manuscripts for publication, books or other materials for review, and correspondence should be addressed to the publishers.

WARWICK AND YORK, INC., *Baltimore, Maryland*

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXIII | September, 1932 | Number 6

A STANDARDIZED TEST OF PROSE APPRECIATION FOR SENIOR HIGH SCHOOL PUPILS

HERBERT A. CARROLL

University of Minnesota

Literature appreciation is a very complex ability, based upon information, sensitivity to style, sympathetic interpretation of fine shades of emotion, intellectual understanding of the deeper meanings of a selection, and an interest in plot complications. But greater than any one of these—perhaps because it includes all or most of them—is the power to differentiate the good from the less good, and the less good from the very bad. The high school student who recognizes the superiority of a passage from Willa Cather over a passage from one of the less literary news-stand magazines possesses discrimination, or the ability to appreciate. It was upon this premise that the writer set about the task of constructing a test of prose appreciation.

SELECTION OF TEST MATERIALS

In choosing selections for inclusion in the test, three factors were considered: Subject, length, and source. If several selections are to be compared with one another, then they should be on the same topic. Seventy-two topics were chosen, and eventually seventy-two sets of selections, four in each set, built.

It was thought desirable, though not absolutely necessary, to have the passages short and approximately equal in length. Those used were taken from four sources: First choices, from books generally considered to be excellent; second choices, from books generally considered to be of poor quality; third choices, from the least literary magazines; fourth choices, from mutilations. All of the above selections were originally chosen with the advice and cooperation of a literary critic. A sample set (No. III in the test) follows:

AN INTERIOR[1]

A

I went with the little maid into a gorgeously decorated bedroom, all of cream color and light blue that blended prettily. The bed was a great, wide affair of beautifully carved and ornamented wood, painted creamy white with blue and gold trimmings. There was a wonderful bureau and a dressing table to match, and in one corner of the room a mirror that went from floor to ceiling. I had to hold my breath.

B

Lollie had never seen such a pretty room, and it made her gasp to see how pretty the furniture was, as well as how pretty the rugs were, and the curtains at the windows and the pictures on the wall, but what she really liked best was that furniture, for it looked comfortable as well as pretty, and she knew it must have cost hundreds and hundreds of dollars. She wished she could live and die in that one room, it was so pretty.

C

An air of Sabbath had descended on the room. The sun shone brightly through the window, spreading a golden lustre over the white walls; only along the north wall, where the bed stood, a half shadow lingered . . . The table had been spread with a white cover; upon it lay the open hymn book, with the page turned down. Beside the hymn book stood a bowl of water; beside that lay a piece of white cloth . . . Kjersti was tending the stove, piling the wood in diligently . . . Sorine sat in the corner, crooning over a tiny bundle; out of the bundle at intervals came faint, wheezy chirrups, like the sounds that rise from a nest of young birds.

D

Major Prime had the west sitting-room. It was lined with low bookcases, full of old, old books. There was a fireplace, a winged chair, a broad couch, a big desk of dark seasoned mahogany, and over the mantel a steel engraving of Robert E. Lee. The low windows at the back looked out upon the wooded green of the ascending hill; at the front was a porch which gave a view of the valley.

VALIDITY

The most difficult requirement which the psychologist must satisfy in constructing an objective measuring instrument of such an intangible variable as appreciation is that his test be valid—that it measure what it purports to measure. The Carroll Prose Appreciation Test depends for its validity upon two criteria: first, source; second, expert opinion. The first of these has already been touched upon. The assumption is made that a selection, representative of the art of Sigrid Undset, would be superior to a typical selection from a wood-pulp magazine, or that a description of spring from Tolstoi could

[1] Key: First, *C*: Rolvaag, O.: "Giants in the Earth." New York: Harper & Bro., 1927, pp. 243–244. Second, *D*: Bailey, T.: "The Trumpeter Swan." Philadelphia: Penn. Pub. Co., p. 26. Third, *A*: Anonymous, Love-starved. *True Experiences*, Macfadden Pub., Nov., 1930, p. 27. Fourth, *B*.

easily be distinguished from one badly written by deliberate intent. To be sure, the passages are short; but a sensitive reader can, even in a few lines, discover an incorrectly used word, an inartistic combination of phrases, or a faulty selection of detail. If only a single note is flatted by a singer, the musician hears it; "errors," if they may be called such, in prose are not so noticeable, but it is at least possible to make gross discriminations.

The writer intends neither to underestimate nor to overestimate the value of source as a criterion of the validity of the test under discussion. In the great majority of instances it would probably be, of itself, sufficient. Certainly, if the only way that a person can tell the difference between the writings of Mr. "X" in "Hot Dog" and Edith Wharton in "Ethan Frome" is by having the two carefully labeled with the name of the author and a statement of the place where the material appears, then that person can hardly be said to possess discrimination in the field of literature. It reminds one of the paintings that have been overlooked because they had not previously been identified as the work of one of the great artists.

However, in a scale of this kind there may be selected occasional passages of very nearly equal value, since at times an inferior author succeeds in writing a few paragraphs that are excellent, while a writer of great merit drops to mediocrity. Largely because of the existence of such a situation, the criterion of source has been supplemented by the criterion of expert opinion in determining the order of merit of the passages in the several sets. Only those selections on which there has been, among the sixty-five judges, a consensus of opinion in absolute agreement with the source are included in the test. (See Table I.)

It will be seen from Table I that the writer made use of three groups of judges; critics, members of University English staffs, and high school teachers of English. Listed as critics are book reviewers, authors of fiction, and librarians. Numbered in this group are men and women internationally known in the world of literature.

Unfortunately, it was impossible to have the same judges for all twelve sets, due principally to the fact that the test was revised several times. A total of twenty-five rendered judgments on all of the selections appearing in the final test; one group of sixty-five judged sets I, III, VI, VIII, IX, and X; another group of sixty-five—with the above-mentioned overlapping—judged sets IV, V, VII, XI, and XII; while a third, made up of members of these two groups, judged set II.

In order to see if the critics, university instructors, and high school teachers would agree, as separate groups, with the criterion of source

TABLE I.—VALIDITY ESTABLISHED BY TWO CRITERIA: SOURCE AND THE JUDGMENTS OF SIXTY-FIVE EXPERTS

Selections	I					II					III					IV					V					VI					Selections
	Critics	University	High school	Total	Source	Critics	University	High school	Total	Source	Critics	University	High school	Total	Source	Critics	University	High school	Total	Source	Critics	University	High school	Total	Source	Critics	University	High school	Total	Source	
A1	5	4	5	14	..	..	..	1	1	..	..	1	..	1	..	6	13	20	39	1	7	22	32	61	1	..	1	1	2	..	A1
2	12	11	13	36	2	..	..	..	..	..	1	2	1	4	..	1	10	12	23	..	..	1	3	4	..	..	1	1	2	..	2
3	3	7	4	14	..	..	..	1	1	..	16	17	19	52	3	..	..	3	3	..	..	..	..	..	..	2	4	1	7	..	3
4	..	..	1	1	..	13	20	30	63	4	3	2	3	8	..	..	..	..	..	..	..	..	..	..	..	18	16	20	54	4	4
B1	8	12	16	36	1	8	8	14	30	1	..	..	1	1	..	..	..	..	..	..	..	1	3	4	..	..	6	1	7	..	B1
2	6	6	6	18	..	2	5	9	16	..	1	..	..	1	..	..	1	4	5	..	6	14	19	39	2	6	4	4	14	..	2
3	6	3	1	10	..	3	7	9	19	..	2	2	3	7	..	7	20	29	56	3	1	8	13	22	..	13	11	16	40	3	3
4	..	1	..	1	..	..	..	..	..	..	17	20	19	56	4	..	2	2	4	..	..	..	..	..	..	1	1	2	4	..	4
C1	7	6	1	14	..	4	9	13	26	..	20	19	23	62	1	1	11	14	26	..	..	..	..	..	..	15	6	3	24	..	C1
2	2	5	4	11	..	8	10	13	31	2	..	3	..	3	..	6	12	19	37	2	..	..	..	..	..	4	12	13	29	2	2
3	10	11	18	39	3	1	1	6	8	..	..	..	..	..	..	..	..	2	2	..	..	..	..	..	..	1	1	6	8	..	3
4	1	..	..	1	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	7	23	35	65	4	..	3	1	4	..	4
D1	..	..	1	1	..	1	3	4	8	..	..	2	..	2	..	..	..	..	..	..	..	..	..	..	..	5	9	18	32	1	D1
2	..	..	..	..	..	3	5	10	18	..	18	17	22	57	2	..	..	..	..	..	1	8	13	22	..	10	5	4	19	..	2
3	1	1	..	2	..	9	12	16	37	3	2	3	1	6	..	..	4	2	6	..	6	15	22	43	3	4	6	1	11	..	3
4	19	21	22	62	4	..	..	2	2	..	..	..	..	..	..	7	19	33	59	4	..	..	..	..	..	1	2	..	3	..	4
A	2	2	2	2	2	4	4	4	4	4	3	3	3	3	3	1	1	1	1	1	1	1	1	1	1	4	4	4	4	4	A
B	1	1	1	1	1	1	1	1	1	1	4	4	4	4	4	3	3	3	3	3	2	2	2	2	2	3	3	3	3	3	B
C	3	3	3	3	3	2	2	T	2	2	1	1	1	1	2	2	2	2	2	2	4	4	4	4	4	1	2	2	2	2	C
D	4	4	4	4	4	3	3	3	3	3	2	2	2	2	2	4	4	4	4	4	3	3	3	3	3	2	1	1	1	1	D

Note: T = tie vote between two selections.

TABLE I.—*Continued*

Selections	VII					VIII					IX					X					XI					XII					Selections
	Critics	University	High school	Total	Source	Critics	University	High school	Total	Source	Critics	University	High school	Total	Source	Critics	University	High schoo	Total	Source	Critics	University	High school	Total	Source	Critics	University	High school	Total	Source	
A1	..	..	..	..	..	18	21	18	57	1	10	10	8	28	..	6	5	6	17	..	..	2	6	8	..	..	1	..	1	..	A1
2	..	..	..	..	..	1	1	3	5	..	9	10	11	30	2	14	13	11	38	2	2	8	4	14	..	2	3	4	9	..	2
3	..	..	3	3	..	1	..	..	1	..	1	2	4	7	..	..	4	6	10	..	4	10	23	37	3	5	17	26	48	3	3
4	7	23	32	62	4	..	..	2	2	..	..	..	..	..	..	..	..	..	..	..	1	3	2	6	..	..	2	5	7	..	4
B1	1	2	1	4	..	1	..	..	1	..	..	..	..	..	..	..	3	3	6	..	2	2	11	15	..	5	20	22	47	1	B1
2	..	3	8	11	..	2	2	2	6	..	1	..	..	1	..	1	2	6	9	..	3	11	16	30	2	2	2	10	14	..	2
3	6	18	24	48	3	3	5	3	11	..	1	1	1	3	..	18	17	10	45	3	2	9	7	18	..	..	1	3	4	..	3
4	..	..	2	2	..	14	15	18	47	4	18	21	22	61	4	1	..	4	5	..	..	1	1	2	..	..	..	..	..	..	4
C1	3	12	27	42	1	..	..	1	1	..	..	2	3	5	..	..	..	..	..	..	5	20	18	43	1	..	..	..	..	..	C1
2	3	7	7	17	..	9	8	8	25	..	1	2	4	7	..	..	..	1	1	..	1	2	13	16	..	..	2	3	5	..	2
3	1	4	1	6	..	10	9	12	31	3	17	17	15	49	3	1	..	4	5	..	1	1	3	5	..	..	2	4	6	..	3
4	..	..	..	..	..	1	5	2	8	..	2	1	1	4	..	19	22	18	59	4	..	..	1	1	..	7	19	28	54	4	4
D1	3	9	7	19	..	1	1	4	6	..	10	10	12	32	1	13	14	15	42	1	..	..	..	..	..	2	2	13	17	..	D1
4	4	13	20	37	2	8	11	11	30	2	9	10	8	27	..	6	7	5	18	..	1	2	2	5	..	3	16	18	37	2	2
3	..	1	6	7	..	6	8	8	22	..	1	2	3	6	..	1	1	3	5	..	..	3	2	5	..	2	3	2	7	..	3
4	..	..	2	2	..	5	2	..	7	..	..	..	..	..	..	..	..	..	..	..	6	18	31	55	4	..	2	2	4	..	4
A	4	4	4	4	4	1	1	1	1	1	1	T	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	A
B	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1	B
C	T	1	1	1	1	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	1	1	1	1	1	4	4	4	4	4	C
D	2	2	2	2	2	2	2	2	2	2	1	T	1	1	1	1	1	1	1	1	4	4	4	4	4	2	2	2	2	2	D

Note: T = tie vote between two selections.

and with one another, their judgments were segregated and tabulated. The results appear in Table I. Even with so great a reduction in numbers, with the attendant increase in possibilities for error and disagreement, there is unanimity among the three groups and source on forty-one of the forty-eight selections. Of the remaining seven, there are four tie votes and only three disagreements. This appears to the writer to be a rather striking correlation.

It will be seen that it is the *consensus* of the experts' judgments that is being accepted. In some instances the discrepancy between an evaluation by a single expert and the criterion of source is as great as an interchange of a first and third choice, or even of a first and fourth. This may be the result of an actual difference of opinion, or, what is more likely, of a hasty rating or even of careless errors in recording decisions.

The validity of the test, then, rests upon the criteria of source and expert opinion. If a high school student agrees, in his judgments, with the source and with the consensus of the rating given by the experts, he is said to be correct—to possess the capacity to appreciate prose literature. If he does not agree, he is said to be wrong.

RELIABILITY

Table II explains itself. It is interesting to note that the two reliability coefficients are the same, *i.e.*, .71. This is much higher,

TABLE II.—RELIABILITY COEFFICIENTS FOR SENIOR HIGH SCHOOL PUPILS

	Re-test	Split-halves
N	467	300
r (with PE)	.71 ± .016	.71 ± .019

insofar as the writer has been able to discover, than that of any other test of literature appreciation. It may be compared, for example, with the .44 published for the Abbot-Trabue Poetry Test.

ORDER OF DIFFICULTY

The twelve sets of selections which make up the total test have been arranged in order of difficulty. Percentages appear in Table III. It will be readily seen that the intervals of difficulty between the several sets are by no means equal. To have made them so would have

required more time and labor than would seem justified by the possible slight increase in reliability and smoothness of the distribution curve.

TABLE III.—ORDER OF DIFFICULTY OF THE TWELVE SETS OF SELECTIONS AS DETERMINED BY PERCENTAGE OF CORRECT SCORES MADE BY THREE HUNDRED HIGH SCHOOL PUPILS

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
58.9	58.1	57.3	53.6	53.3	52.6	47.1	43.7	42.4	41.0	37.8	35.8

ADMINISTRATION AND SCORING

Detailed directions for the administration and scoring of the test appear in the Manual. Only two points need be considered here.

First, there is no time limit, as it was felt that it would be unwise to try to put a stop-watch on appreciation. Nearly all high school students will finish within the usual forty minute period. When the test was standardized only one hundred eighty-seven pupils in over three thousand handed in incomplete record sheets, the time allowed being a single period of forty minutes. In order to learn whether or not these one hundred eighty-seven cases would have seriously affected the final distribution of scores, they were matched on the basis of Grade and School with one hundred eighty-seven children, chosen at random, who had completed the test. The papers were then scored on the basis of the first half only, and the means compared. (See Table IV.)

TABLE IV.—A COMPARISON OF THE DISTRIBUTION OF SCORES ON THE FIRST HALF OF COMPLETE AND INCOMPLETE RECORD SHEETS

	Completed records	Incompleted records
N	187	187
M (with PE)	19.08 ± .307	18.12 ± .312
σ	6.22	6.32

$M_{Com.} - M_{Incom.} = 19.08 - 18.12 = .96 \pm .438$

Since the difference between the means divided by its probable error is only a little over two, it was concluded that these incomplete records could be thrown out without the final distribution being appreciably affected.

The second point concerns scoring. Several methods were experimented with, the one giving the highest reliability being that of weighting the first and fourth choices. Viewed subjectively, as well, this seemed to be the most satisfactory. The method used, then, places a premium upon the ability to recognize the very best and the very worst; or, to put the latter part of the above statement negatively, penalizes an individual for inability to recognize the almost complete lack of merit in the mutilations. Numerically, two points each are given for the correct ranking of selections I and IV, and one point each for the correct ranking of selections II and III.

DISTRIBUTION OF SCORES

In its preliminary and final forms the Carroll Prose Appreciation Test has been given to approximately nine thousand people of varying ages and educational levels. It is a product, then, of a series of trials on actual school populations. In all, seventy-two sets were used at one time or another; of these, twelve remain. This final form has been standardized for Grades X, XI, and XII. The subjects used were children in the four following high schools: Mechanic Arts, Central, and Humboldt, all of St. Paul, Minnesota, and the University High School

TABLE V.—DISTRIBUTION OF SCORES MADE BY SENIOR HIGH SCHOOL PUPILS

Scores	Grade X	Grade XI	Grade XII	Total
60–63	0	1	2	3
56–59	5	6	17	28
52–55	10	30	44	84
48–51	33	54	90	177
44–47	66	104	118	288
40–43	133	134	150	417
36–39	153	177	171	501
32–37	143	148	153	444
28–31	139	126	108	373
24–27	117	97	76	290
20–23	105	63	37	205
16–19	59	40	23	122
12–15	27	14	7	48
8–11	8	5	3	16
4– 7	1	1	1	3
0– 3	1	0	0	1
N	1000	1000	1000	3000

at the University of Minnesota. The distribution of the scores made by three thousand of these pupils appears in Table V.

In Table VI appear data relative to the central tendency and variability of the scores made by several groups.

TABLE VI.—MEASURES INDICATING THE DISTRIBUTION OF THE SCORES MADE BY THE SEVERAL GROUPS THAT TOOK THE FINAL FORM OF THE CARROLL PROSE APPRECIATION TEST

Measures	High school pupils				College students	Judges
	Grade X	Grade XI	Grade XII	Total		
N	1000	1000	1000	3000	225	25
H. P. S	72	72	72	72	72	72
Range	1–59	6–60	5–63	1–63	7–65	34–64
M (with PE)	32.83	35.56	38.00	35.45	43.32	53.00
	±.204	±.202	±.201	±.120	±.425	±.707
σ	9.56	9.48	9.41	9.72	9.45	5.24
M. (with PE)	33.22	36.18	38.32	35.98	43.55	54.33
	±.255	±.252	±.251	±.150	±.531	±.884
Q_3	39.90	42.40	44.58	42.38	50.11	57.25
Q_1	25.41	28.94	31.79	28.69	37.34	49.25
Q	7.24	6.73	6.40	6.84	6.38	4.00

Besides presenting facts concerning the distribution of the scores, this table gives additional proof of the validity of the test. It will be noted that there is a steady increase in central tendency from Grade X, through college, to the judges. Table VII shows clearly that the existing difference between the means of any two groups is a highly reliable one. The consistency of the differences with respect to high school students is shown in Figure 1.

TABLE VII.—DIFFERENCES, WITH THEIR PROBABLE ERRORS, BETWEEN CERTAIN PAIRS OF MEANS

$$M_{(Judges)} - M_{(Col.\ Students)} = 53.00 - 43.32 = 9.68 \pm .825$$
$$M_{(Col.\ St.)} - M_{(H.\ S.\ Stud.)} = 43.32 - 35.45 = 7.87 \pm .442$$
$$M_{(Gr.\ XII)} - M_{(Gr.\ XI)} = 38.00 - 35.56 = 2.44 \pm .285$$
$$M_{(Gr.\ XI)} - M_{(Gr.\ X)} = 35.56 - 32.83 = 2.73 \pm .287$$

SUMMARY

The Carroll Prose Appreciation Test for senior high school students is based upon the assumption that the ability to appreciate literature

can be measured by revealing the degree to which an individual discriminates among passages of varying worth. Its validity has been determined by two criteria: source and a consensus of the judgments of

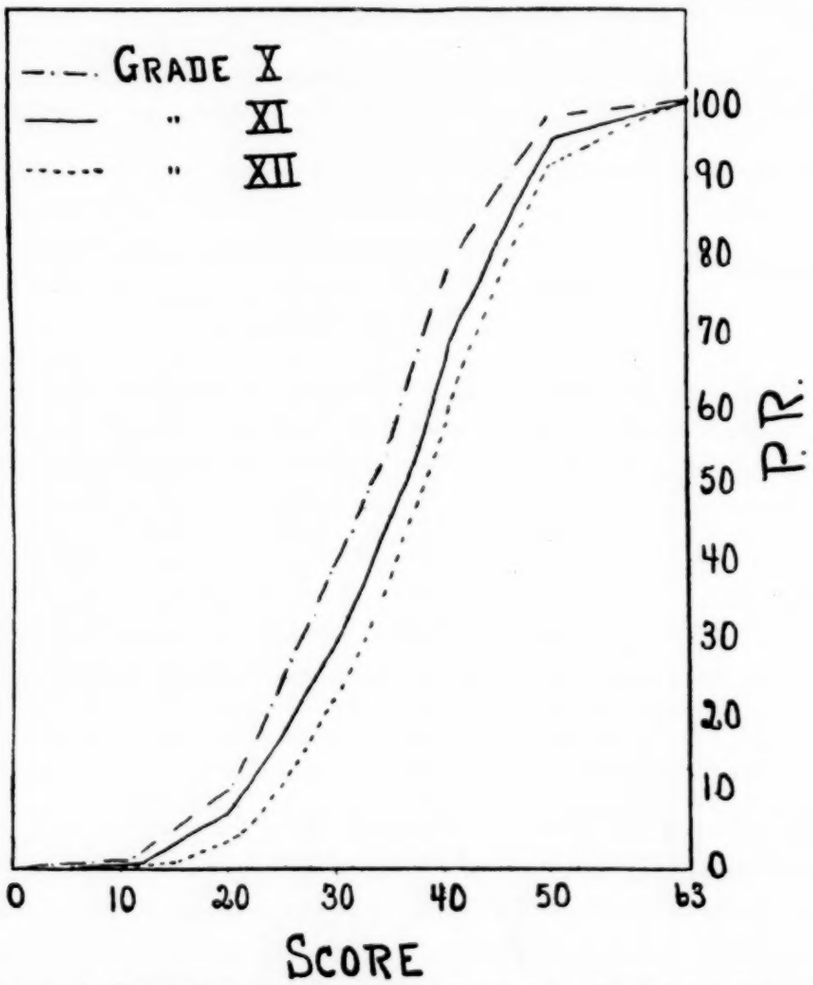


FIG. 1.—Percentile Distribution of Scores Made on the Carroll Prose Appreciation Test by 1000 Children from Each of the Senior High School Grades.

sixty-five experts, the correlation between the two being perfect. Its reliability, found for four hundred sixty-seven senior high school students, is .71 ± .016. The amount of time needed for taking the test is approximately thirty-five minutes. Tentative norms have been established on three thousand high school pupils. The publisher is: Educational Test Bureau, Minneapolis, Minnesota.

THE RELIABILITY OF A SINGLE TEST ITEM

KARL J. HOLZINGER

The University of Chicago

INTRODUCTION

In this paper we shall introduce formulae giving the standard error of response or measurement of a single test item. The response error of series of items is already known. The problem, then, is to find this error for a single item and show the relation to the standard error of a series of such items.

It is hoped that these new formulae will be of some value in building up tests of items of known reliability and in predicting the final reliability of a number of such items when combined. Conversely it is believed the formulae may be useful in appraising the reliability of tests of different lengths by reducing the measure of reliability to the average item basis.

THE ERROR OF RESPONSE OR MEASUREMENT OF A LENGTHENED TEST

Before introducing the new formulae we shall examine the effect of test lengthening upon the usual formula

$$\sigma_e = \sigma_1\sqrt{1 - r_{11}} \tag{1}$$

This formulae may be derived in several ways, one of which is as follows:

Let x_1 and x_1 = scores on two forms of a test, expressed as deviations from respective means,

r_{11} = reliability coefficient of these forms

σ_1 and σ_1 = standard deviations of such scores

x_∞ = "true" score on either form (average of large number of testings)

e_1 and e_1 = uncorrelated errors defined by the expressions

$$x_1 = x_\infty + e_1$$
$$x_1 = x_\infty + e_1$$

Subtracting, member by member, gives

$$x_1 - x_1 = e_1 - e_1$$

Squaring, summing, and dividing by N (number of cases) gives

$$\sigma_1^2 - 2r_{11}\sigma_1\sigma_1 + \sigma_1^2 = \sigma^2_{e_1} + \sigma^2_{e_1}$$

If $\sigma_1 = \sigma_I$ and $\sigma_{e_1} = \sigma_{e_I}$ we then have

$$\sigma_e^2 = \sigma_1^2(1 - r_{II}) \tag{2}$$

σ_e is thus the standard error of response of $(x_1 - x_\infty)$.

We shall inquire next into the effect of test lengthening upon formula (2). The reliability of a lengthened test is given by the well-known Spearman-Brown formula

$$r_{nn} = \frac{nr_{II}}{1 + (n - 1)r_{II}} \tag{3}$$

The standard deviation of a lengthened test is expressed by the easily derived formula

$$\sigma_n = \sigma_1\sqrt{n + n(n - 1)r_{II}} \tag{4}$$

Writing the standard error of response of the lengthened test as

$${}_n\sigma_e = \sigma_n\sqrt{1 - r_{nn}}$$

we have, upon substituting formulae (3) and (4),

$$\frac{{}_n\sigma_e}{\sqrt{n}} = \sigma_1\sqrt{1 - r_{II}} = \sigma_e \tag{5}$$

In the derivation of the last three formulae it has been assumed that the intercorrelations of the component tests are all equal, whether on a single form or between any two components on similar forms of the tests. Thus if the components on each form be designated as ${}^1x_1$, ${}^2x_1$, ${}^3x_1$, . . . ${}^nx_1$ and ${}^1x_I$, ${}^2x_I$, ${}^3x_I$, . . . ${}^nx_I$ respectively, all of the n^2 intercorrelations, $\Sigma^i x_1{}^i x_I/N\sigma_1\sigma_I$ $(i = 1$ to $n)$ are assumed equal. The standard deviations of ${}^ix_1$ and ${}^ix_I$ are also assumed equal.

If two forms of a test have been prepared so that component ${}^1x_1$ is very similar to ${}^1x_I$, component ${}^2x_1$ is very similar to ${}^2x_I$ etc., then the correlations between ${}^1x_1$ and ${}^1x_I$ or between ${}^2x_1$ and ${}^2x_I$ will usually be higher than the cross values between ${}^1x_1$ and ${}^2x_I$ or between ${}^2x_1$ and ${}^1x_I$, etc. We shall therefore distinguish two types of such coefficients, *viz.*, n parallel values ${}_pr_{II}$ between ${}^1x_1$ and ${}^1x_I$ or ${}^2x_1$ and ${}^2x_I$, etc., and $n(n - 1)$ cross values designated as ${}_cr_{II}$ between ${}^1x_1$ and ${}^2x_I$ or between ${}^2x_1$ and ${}^1x_I$ etc. With this distinction formula (3) would be written

$$r_{nn} = \frac{n_pr_{II} - n(n - 1)_cr_{II}}{n + n(n - 1)_cr_{II}} \tag{3'}$$

Formula (4) would be expressed as

$$\sigma_n = \sigma_1\sqrt{n + n(n-1)_c r_{11}} \tag{4'}$$

and formula (5) would become

$$\frac{{}_n\sigma_e}{\sqrt{n}} = \sigma_1\sqrt{1 - {}_p r_{11}} \tag{5'}$$

Formula (5) is important because it shows that the response error is a function of the length of a test. We cannot, therefore, make adequate comparison of the response error in tests until their varying length has been taken into account. Thus if a test of 25 items has a response error of σ_e, this error will be increased to $2\sigma_e$ upon increasing the test to one hundred items, etc.

Formula (5′) shows that in obtaining the response error of a lengthened test we need consider only the direct component reliability coefficients ${}_p r_{11}$ of which there are n in number. This greatly simplifies the problem because all of the $n(n-1)$ cross coefficients may be ignored in obtaining the value ${}_n\sigma_e$.

If the reliability of a lengthened test is required formula (3′) will give much more accurate results than formula (3) if the parallel and direct coefficients differ greatly in value (as they often do). The component standard deviations should also be taken into account, but the complexity of the resulting formulae makes this unpractical. We shall therefore assume these all equal with a certain undetermined error resulting.

THE ERROR OF RESPONSE OF A SINGLE ITEM

In order to obtain an expression for σ_e which will apply to a pair of equivalent items we shall first express formula (1) in a different form: Let

$$d = x_1 - x_{\text{I}}$$

and

$D = X_1 - X_{\text{I}}$ for raw scores with means M_1 and M_{I}.

From these last expressions we find that

$$\sigma_d{}^2 = 2\sigma_1{}^2(1 - r_{11}) = 2\sigma_e{}^2 \tag{6}$$

and

$$\sigma_d{}^2 = \frac{\Sigma D^2}{N} - (M_{\text{I}} - M_1)^2 \tag{7}$$

Therefore

$$\sigma_e^2 = \frac{1}{2}\left[\frac{\Sigma D^2}{N} - (M_{\text{I}} - M_1)^2\right] \tag{8}$$

Formula (8) will next be applied to the case of a pair of single items, one for each form of a test, and each scored either 0 or 1. If N pupils try both of these items, the scores on the pair of items are recorded as (0, 1), (1, 1) (0, 0) or (1, 0). The number of pupils getting these four possible pairs of scores will be denoted by a, b, c, and d respectively as illustrated by the following scheme:

Score on Form X_{I}	a	b
	c	d

Score on Form X_1

Now it is evident that $D = X_1 - X_{\text{I}}$ can take only the four values $(0, -1)$, $(1, -1)$, $(0, -0)$ and $(1, -0)$ so that $\Sigma D^2 = a + d$. It is also apparent that $M_{\text{I}} - M_1$ for the above table $= \dfrac{d - a}{N}$

We thus find

$${}_1\sigma_e^2 = \frac{1}{2}\left[\frac{a + d}{N} - \left(\frac{a - d}{N}\right)^2\right] \tag{9}$$

as the desired standard error for a pair of items. If $a = d$ this formula reduces to

$${}_1\sigma_e^2 = \frac{a}{N} \tag{10}$$

If $(a - d)^2/N$ is negligible but $a - d \neq 0$ we may write

$${}_1\sigma_e^2 = \frac{1}{2}\left[\frac{a + d}{N}\right] \tag{11}$$

The reliability coefficient for the pair of items may also be found from the point correlation coefficient

$${}_1r_{11} = \frac{bc - ad}{\sqrt{(a + b)(c + d)(a + c)(b + d)}} \tag{12}$$

and the standard deviations are given by

$${}_1\sigma_1 = \frac{\sqrt{(a + c)(b + d)}}{\sqrt{N}} \tag{13}$$

$${}_1\sigma_{\text{I}} = \frac{\sqrt{(a + b)(c + d)}}{\sqrt{N}} \tag{14}$$

These two standard deviations have been assumed equal in formula (9), so that $b = c$. The correlation (12) may then be written

$$_1r_{11} = \frac{b^2 - ad}{(a + b)(b + d)} \tag{15}$$

If $a = d$ this formula is further reduced to

$$_1r_{11} = \frac{b - a}{b + a} \tag{16}$$

As a check on the appropriateness of the point correlation for these purposes we may solve the expression $\sigma_e^2 = \sigma_1^2 (1 - r_{11})$ or

$$\frac{1}{2}\left[\frac{a + d}{N} - \left(\frac{a - d}{N}\right)^2\right] = \frac{(a + b)(c + d)}{N^2}(1 - r_{11}) \text{ for } r_{11}$$

The solution gives formula (15).

THE RELATION OF THE ERROR OF A PAIR OF ITEMS TO THAT OF A LONG TEST

Returning to the formula

$$\frac{_n\sigma_e}{\sqrt{n}} = \sigma_1\sqrt{1 - {}_pr_{11}} \tag{5'}$$

we may now regard n as the number of items in the test and write

$$\frac{_n\sigma_e}{\sqrt{n}} = \frac{1}{2}\left[\frac{a + d}{N} - \left(\frac{a - d}{N}\right)^2\right] \doteq \frac{1}{2}\left[\frac{a + d}{N}\right] \tag{17}$$

This formula may be used to calculate the error of estimate of a long test composed of items the error of which has been found by formula (9). If the errors of the individual items are all equal the formula is exact; if the errors are not equal their average may be taken as a good approximation in obtaining a value for $_n\sigma_e$ as will be shown below.

As a theoretical illustration assume one hundred test item pairs each of response error 0.2 as determined by formula (9). The standard error of response of a test of one hundred such items (two forms one hundred items each) would be predicted by formula (17) to be two.

Conversely, if a test has a response error of two and is composed of one hundred items, formula (5′) shows that the standard error of each item pair is 0.2 (assuming these latter errors equal). It will be shown below that formula (5′) gives a good measure of the standard error of the average test item pair when these errors vary considerably.

Two suggested uses of the above formulae are:

1. Given the standard error of response of long tests composed of items scorable 0 or 1 formula (5′) gives a good measure of the average response error of test item pairs. This makes possible the direct comparison of the response errors in tests on the item basis. Thus if one test furnished a value from formula (5′) of 0.2, while another test a value of 0.4, we can say the former is composed of better items in the sense that their average response error is less.

2. In building a test, two forms may be prepared and given to a group of pupils. The values of ${}_1\sigma_e$ may then be calculated from formula (9) or (11) and a group of pair items selected for the long test on the basis of small average response error. From these chosen items the standard error of the proposed long test may be calculated by formula (5′) or (17). Permissible variations in ${}_1\sigma_e{}^2$ are discussed below.

THE SAMPLING ERROR OF ${}_1\sigma_e{}^2$

Formula (11) indicates that ${}_1\sigma_e$ is approximately a proportion, so that its standard error of sampling is given by the formula

$$\text{SD (sampling of } {}_1\sigma_e{}^2) \doteq \sqrt{\frac{{}_1\sigma_e{}^2(1 - {}_1\sigma_e{}^2)}{N}} \tag{18}$$

It is now suggested that in building a test as described in the preceding section, variations of ${}_1\sigma_e{}^2$ about the average $\sigma_e{}^2$ be used within a range of ± 3SD (sampling of the average ${}_1\sigma_e{}^2$). Experimental verification of this proposal is given in the next section.

We shall take as an approximation to the standard error of the average value of ${}_1\sigma_e{}^2$ the quantity

$$\text{SD (sampling error of } \overline{\sigma_e{}^2}) \doteq \sqrt{\frac{\overline{\sigma_e{}^2}(1 - \overline{\sigma_e{}^2})}{Nn}} \tag{19}$$

where $\overline{\sigma_e{}^2}$ denotes the average value.

EXPERIMENTAL VERIFICATION

A number of the writer's students are trying out test material with the purpose of verifying the above formulae. One of these experiments is briefly stated below:

Two forms of Beauchamp's Science Tests of fifty items each were given to one hundred five pupils. All items were scored 0 or 1 and fifty values $\frac{1}{2}[(a + d)/N]$ calculated. The average value of ${}_1\sigma_e{}^2$

from formula (11) was 0.15 ±0.005. The reliability of the fifty items was 0.837 with $\sigma_n = 7.84$ which gives a value of 0.200 when substituted in the formula $\sigma_n^2(1 - r_{nn})/n$. The difference between 0.20 and 0.15 ± 0.005 indicates that the discrepancy exceeds sampling variation and that formula (17) may not be applied to such data.

The discrepancy is probably due to the wide variation in values of $\frac{1}{2}[(a + d)/N]$. To test this hypothesis we next selected twenty-two items with reliability exceeding 0.35 by formula (12) and not differing from one another in sampling error by more than twice the value from formula (18). For these twenty-two pairs of items the average value of ${}_1\sigma_e^2$ from formula (11) was 0.13 ± 0.01. The reliability of the twenty-two items was 0.813 with $\sigma_n = 4.01$ which gives a value of 0.14 when substituted in the formula $\sigma_n^2(1 - r_{nn})/n$. The difference 0.14 − 0.13 ± 0.01 here indicates that formula (17) applies very well to the data.

The suggestions from this experiment are that the assumption underlying formula (17) as applied to average values of $\frac{1}{2}[(a + d)/N]$ are met if variations do not exceed two standard deviations by formula (18), and if items of highest reliability are selected.

The above twenty-two pairs of items were also used to illustrate the use of formula (3) and (3′). The average reliability of these twenty-two items was 0.412. If this value be substituted in formula (3) for $n = 22$ we find $r_{nn} = 0.939 \pm 0.019$ (Shen's formula). This prediction exceeds the observed reliability coefficient 0.813 by an amount beyond the bounds of sampling error. The formula does not apply because the cross correlations were assumed equal to the parallel values. When the former are calculated and used in formula (3′) we find $r_{nn} = 0.847 \pm 0.022$ which does not differ significantly from the observed value 0.813. Clearly formula (3′) should be used to estimate the reliability of a lengthened test where parallel and cross coefficients differ. In this case the average of the parallel coefficients ${}_pr_{11}$ was 0.412, while the average of a sampling of the $n(n - 1)$ cross coefficients ${}_cr_{11}$ was only 0.158.

The length and detail of this paper make it desirable to defer further illustrations until later. It is hoped that those interested in test construction will also check these formulae with experimental data. The writer will also appreciate criticism because he is well aware of several doubtful points in the application of the formulae.

A METHOD FOR DETERMINING THE ABSOLUTE AFFECTIVE VALUE OF A SERIES OF STIMULUS SITUATIONS

PAUL HORST

Personnel Research Division, United States Civil Service Commission

I. INTRODUCTION

The object of this investigation was to develop a technique whereby a series of stimulus situations could be scaled with reference to their affective value. By "affective value" we mean simply the degree to which the stimulus situation evokes feelings of pleasantness or unpleasantness. The method assumes that there is a continuous gradation from those experiences which are extremely unpleasant to those which are extremely pleasant.

For example, let us imagine a scale as indicated in Fig. 1:

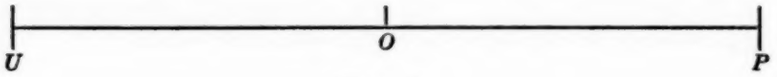


FIG. 1.

We assume that any stimulus situation may be allocated on this scale according to the degree of affect which it normally evokes in an individual. In the following analysis we attempt to measure the distance which any given situation has when allocated with reference to its affective value, from the point of zero affect. This distance is to be measured in terms of units to be subsequently defined.

The statistical technique involved is a modification and extension of a method developed by Thurstone.[1] The technique makes possible the determination of scale separations, on a psychological continuum of items in a series of dissimilar stimulus situations. So far as the method is concerned, these situations may be anything, providing they can be compared with reference to the degree of affect which they evoke.

Evidently any given stimulus situation will not evoke the same degree of affect in all individuals, nor will it evoke the same degree of affect in a given individual from one time to another. Without committing ourselves as to the manner in which an individual evaluates the affective value of a stimulus situation or distinguishes between the relative affective values of two or more situations, we may assume that

some one affective value is associated more frequently with a given situation than is any other value.

For convenience we employ the terminology of Thurstone. He defines the "discriminal process" as that process by which the organism indentifies, distinguishes, discriminates, or reacts to stimuli. The fluctuation in the discriminal process for a uniform repeated stimulus is designated the "discriminal dispersion." The discriminal process which is associated more frequently with a given stimulus situation than with any other situation is termed the "modal discriminal process" for that stimulus situation. Thurstone shows that the fundamental psychophysical equation for determining the scale separation between two stimulus situations is

$$S_i - S_j = b_{i-j}\sqrt{\sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j} \qquad (1)$$

where

S_i is the modal scale value on the psychological continuum for the stimulus situation i.

S_j is the modal scale value on the psychological continuum for the stimulus situation j.

b_{i-j} is the sigma value as found in a table of the normal probability integral for the experimentally observed proportion of judgments "i less than j."

σ_i is the standard deviation of the discriminal dispersion of situation i on the psychological continuum.

σ_j is the standard deviation of the discriminal dispersion of situation j on the psychological continuum.

r_{ij} is the correlation between the discriminal deviations for situations i and j.

The assumptions upon which the validity of the above equation rests are given as follows:

(1) Every stimulus in the stimulus-series is associated with a modal discriminal process with which the organism identifies the stimulus for a prescribed attribute.

(2) The modal discriminal process for any given stimulus retains at least some of its identity even when the stimulus is combined with other stimuli into a single perceptual judgment.

(3) The modal processes may be arranged in a linear psychological continuum in the same serial or rank order as the corresponding stimulus series.

(4) In addition to arranging the discriminal processes in rank or serial order, linear separations between them are assigned on the

assumption that the discriminal dispersion for any stimulus is normal on the psychological continuum. This assumption is subject to experimental verification.

The meaning of equation (1) may be indicated diagrammatically by Fig. 2:

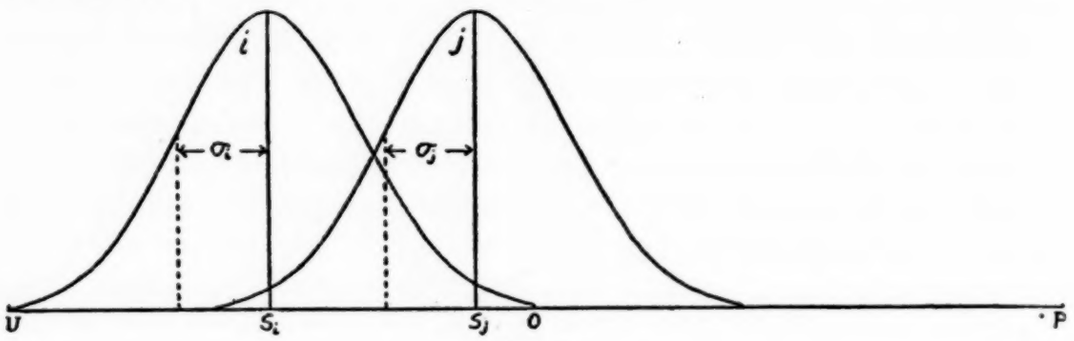


FIG. 2.

The base line of this figure is the affective scale shown in Fig. 1. We have given two situations i and j. For each situation we have a number of different discriminal processes identifying the situation with respect to its affective value. These processes may differ because they come from different individuals, or because they came at different times from the same individual, or both conditions may contribute to the variation. In any event, we obtain a distribution of discriminal processes on the affective scale or continuum for both situation i and j. According to assumption (4) the discriminal dispersions of the two situations respectively are normal. The mean or modal scale value S_i is by definition the affective value of situation i. Similarly S_j is the scale value of situation j. The standard deviations of the discriminal dispersions of the two situations are σ_i and σ_j respectively.

From equation (1) it is clear that the value b_{i-j} is a function both of the distance between S_i and S_j and the dispersions of the two distributions.

If, in addition to the four assumptions previously made we assume that:

(5) The correlation between the discriminal deviations for any two items in the series is zero.

(6) All of the discriminal dispersions in the series are equal. Then equation (1) simplifies to

$$S_i - S_j = b_{i-j}\sigma\sqrt{2} \qquad (2)$$

Thus far no unit has been indicated for measuring the distances on

the affective scale. If we adopt σ, or the standard deviation of the discriminal dispersion, as the unit of measurement, then equation (2) becomes

$$S_i - S_j = b_{i-j}\sqrt{2} \tag{3}$$

It should be noted, however, that equation (3) gives only the distance between S_i and S_j. It does not give the distance of either S_i or S_j from the origin of the scale, that is, from the point of zero affect.

II. THEORETICAL

1. *The Difference Equation.*—We shall now derive equation (3) analytically.

In Fig. 2, S_i has been defined as the modal scale value on the psychological continuum for situation i. Since the discriminal dispersion is, by assumption, normal on the continuum, the equation to the distribution curve is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-s_i)^2}{2}} \tag{4}$$

where

x is the affective value corresponding to a given discriminal process measured from the point of zero affect in units of the discriminal dispersion.

S_i is the modal scale value measured from the point of zero affect in units of the standard deviation of the discriminal dispersion.

Analytically, S_i is that value which in the equation

$$P = \int_{-\infty}^{x-S_i} f(x)dx \tag{5}$$

makes $P = 0.5$.

Since equation (5) is merely the expression for the proportion of the area of the curve below the ordinate erected at x, it is obvious that $P = 0.5$ when $x = S_i$.

Suppose we represent the distribution of discriminal processes associated with situation i by Fig. 3. Then the probability that a discriminal process corresponding to a value below x will be associated with situation i is given by the proportion of the area below the ordinate at x. If this proportion is given we can substitute it in equation (5) and find the corresponding value of $(x - S_i)$ from a normal probability table. Designating this value by (d) we have

$$x - d = S_i \tag{6}$$

where both x and d are given. Hence S_i can be determined.

Theoretically, then, we might determine the affective value of stimulus situation i by selecting any arbitrary value of x and obtaining from a large group of individuals the proportion of the group which judges the affective value of situation i to be less than x. Substituting this proportion in equation (5), (d) in (6) could be determined and S_i solved for.

Obviously, however, the procedure is not feasible since it would be impossible for individuals to agree on the meaning, in terms of affective value, of any given numerical value. Therefore, instead of dealing with distributions of absolute processes, we deal with distributions of

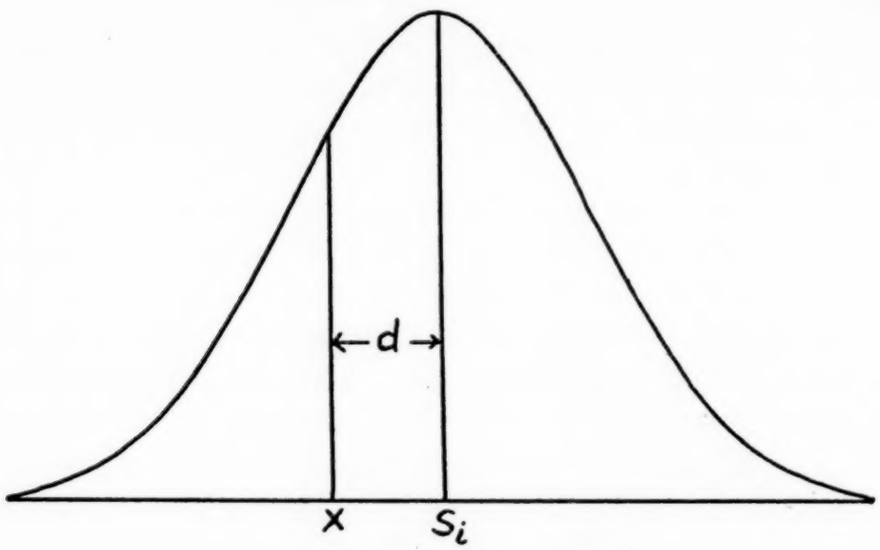


FIG. 3.

differential processes. By an "absolute process" we mean a discriminal process which identifies a single stimulus, whereas, the term "differential" is used to designate a discriminal process which differentiates between two stimuli.

We let x_i and x_j represent the corresponding affective values of the absolute discriminal processes by which a given individual identifies situations i and j, and x_{i-j} the affective value corresponding to the differential process by which he differentiates them.

If the distributions of x_i and x_j are normal, then the distribution of x_{i-j} is also normal. Hence, instead of representing two separate distributions of absolute processes on the psychological continuum, as in Fig. 2, we postulate a single distribution of differential discriminal processes. This is represented diagrammatically in Fig. 4.

Now $S_i - S_j$ is defined as that value for which the probability is 0.5 that a differential discriminal process associated with situations i

and j will lie below $S_i - S_j$. Similarly, the probability is also 0.5 that the process will lie above $S_i - S_j$. Hence $S_i - S_j$ is that value which makes $P = 0.5$ in the equation

$$P_{i-j} = \int_{-\infty}^{x_{i-j}-(S_i-S_j)} f(x_{i-j})dx_{i-j} \tag{7}$$

In general, if we have given the probability that a differential process associated with situations i and j will lie below any value x_{i-j}, we can find the distance between the value x_{i-j} and $S_i - S_j$. The probability that a differential process corresponding to a value

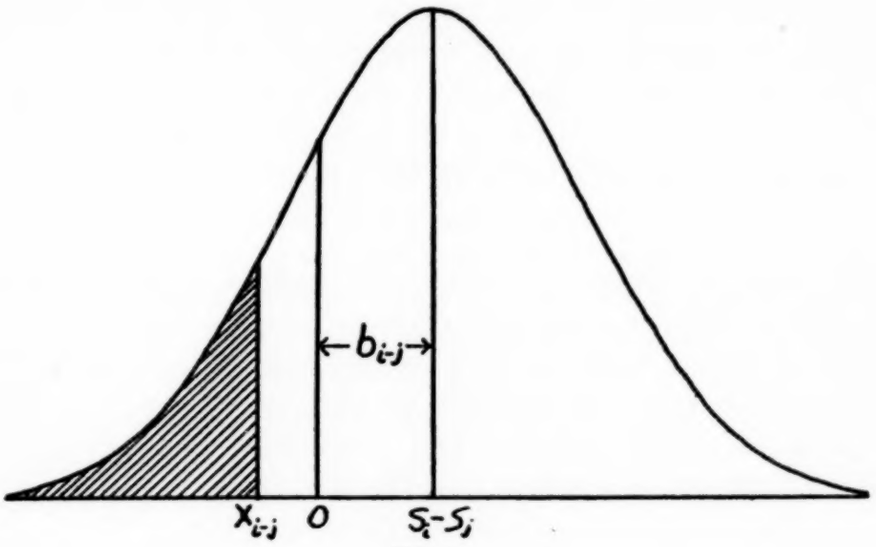


FIG. 4.

below x_{i-j} will be associated simultaneously with situations i and j is given by the proportion of the area of the curve below the ordinate at x_{i-j}. This area is represented by the cross-hatched portion of Fig. 4. If this proportion is given we can substitute it in equation (7) and find the corresponding value of $x_{i-j} - (S_i - S_j)$ from a normal probability table. Designating this value by b_{i-j}, we get

$$x_{i-j} - (S_i - S_j) = b_{i-j}\sqrt{2} \tag{8}$$

where the value $\sqrt{2}$ is the standard deviation of the distribution of differential processes in units of the discriminal dispersion of a distribution of absolute processes.

If we obtain from a large group of individuals the proportion of the group who judge the affective value of situation i to be less than

situation j this means that they judge $S_i - S_j$ to be less than zero. In this case $x_{i-j} = 0$. Substituting this value of x_{i-j} in equation (8) and remembering that b_{i-j} is negative when $S_i < S_j$ and positive when $S_i > S_j$ we get

$$S_i - S_j = b_{i-j}\sqrt{2} \tag{9}$$

which is the same as equation (3). The value b_{i-j} is represented in Fig. 4 by the distance between 0 and the mean $S_i - S_j$.

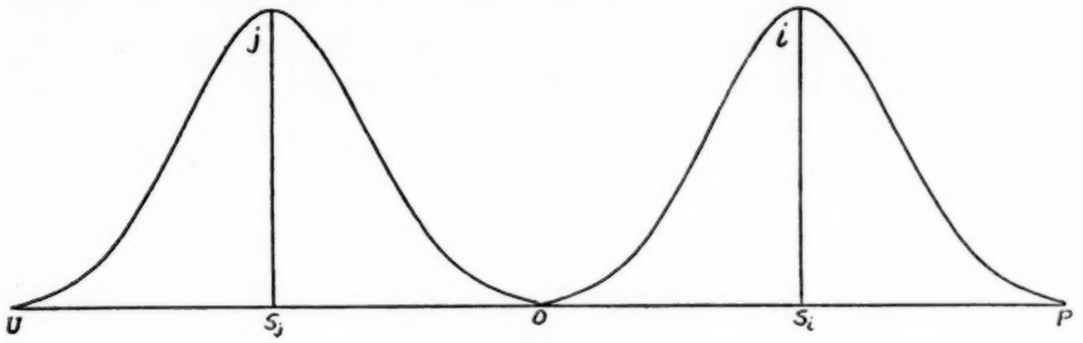


FIG. 5.

2. *The Sum Equation.*—It is evident from the preceding section that in order to find the scale separation between two stimulus situations these situations must be close enough together on the psychological continuum so that one will not be judged greater than the other by all members of the group. The possibility of determining scale separations depends on the fact that there is disagreement as to which of two stimulus situations has the greater value. This principle

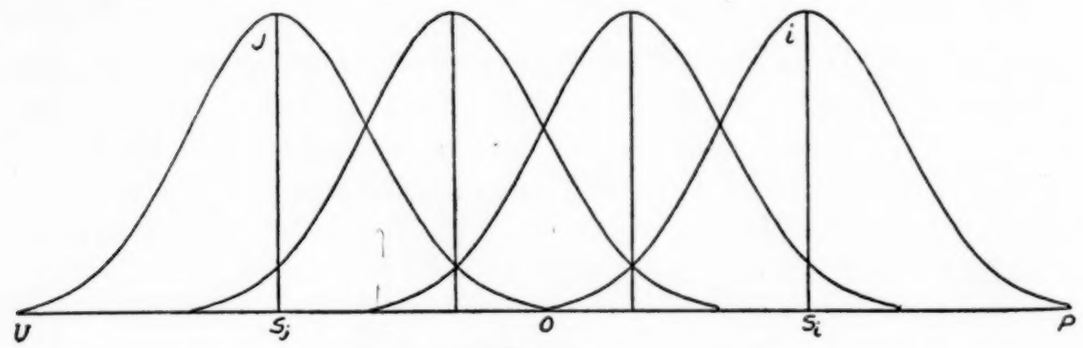


FIG. 6.

has been represented by the overlapping distributions of discriminal processes in Fig. 2.

But suppose we have a situation as indicated in Fig. 5.

Here again we have the same affective scale as in Figs. 1 and 2. The meaning of the distributions is the same as in Fig. 2. But here j

is a very unpleasant situation and i is a very pleasant situation. It is highly improbable that anyone would regard j as a more pleasant situation than i. Since there is not this disagreement, the method of the preceding section cannot be employed to find the distance between S_i and S_j. If, however, we had a number of situations which could be represented by a series of overlapping distributions along the entire range between S_i and S_j, as indicated by Fig. 6,
then the method of the preceding section would be available. We could determine the scale distance between the means of each adjacent pair of distributions and by adding all of these obtain the distance between S_i and S_j.

As a matter of fact, this is a part of the procedure actually employed. But, as will be shown later, this procedure if employed alone enables us to obtain only scale separations, whereas we are primarily interested in finding the distance of each scale value from the origin. For this reason, we develop a method for determining the sum of the scale values of two situations.

First we define a "summative process" as that process by which the organism identifies the summative value of two stimulus situations. We let x_{i+j} represent the affective value corresponding to the summative process by which the individual integrates situations i and j.

Retaining the assumptions of the previous section, we can write an equation analogous to (7), namely,

$$P_{i+j} = \int_{-\infty}^{x_{i+j}-(S_i+S_j)} f(x_{i+j})dx_{i+j} \tag{10}$$

where $S_i + S_j$ is the algebraic sum of the values S_i and S_j, and is defined as that value which makes $P = 0.5$.

To determine this value experimentally, the same logic is employed as in the previous section. We obtain the proportion of individuals who judge the summative value of situations i and j to be less than zero. In equation (10) the value of x_{i+j} in the upper limit of the integral is therefore zero. Since P is given, the upper limit $-(S_i + S_j)$ can be obtained from a normal probability table. Indicating this value by b_{i+j} and multiplying by $\sqrt{2}$, the sigma of the distribution, we get

$$S_i + S_j = b_{i+j}\sqrt{2} \tag{11}$$

remembering that b_{i+j} is negative when $P < 0.5$ and positive when $P > 0.5$.

III. TECHNIQUE

1. *Calculation of Scale Separations by the Average Difference Method.* The method employed by Thurstone[1] in solving for the scale separations of stimulus values is as follows:

Having given the observation equations of the type

$$S_i - S_j = b_{i-j}\sqrt{2}$$

the constant values b_{i-j} are arranged in a skew symmetrical table thus:

	1	2	3	4		n
1	0	$-b_{1-2}$	$-b_{1-3}$	$-b_{1-4}$	. . .	$-b_{1-n}$
2	b_{1-2}	0	$-b_{2-3}$	$-b_{2-4}$	. . .	$-b_{2-n}$
3	b_{1-3}	b_{2-3}	0	$-b_{3-4}$	. . .	$-b_{3-n}$
.	.	.	.	.	.	.
n	b_{1-n}	b_{2-n}	b_{3-n}	b_{4-n}	. . .	0

(12)

where

b_{i-j} is the sigma value of the proportion of individuals who judge $S_i - S_j$ to be less than 0, or, what amounts to the same thing, who judge S_i to be less than S_j.

It will be noted that the entry in the i'th row of the j'th column is always the same as the entry in the j'th row of the i'th column with sign changed. That is $-b_{ij} = +b_{ji}$. Only the values below the diagonal are given in the observation equations. Those above are copied from those below.

A table of differences is derived from the first table by subtracting column 2 from 1, 3 from 2, 4 from 3, etc. Each remainder in a given difference column is of the type $b_{i(K-1)}$, $- b_{iK}$ so that summing for any given column and multiplying by $\sqrt{2}/n$ we get

$$S_{K-1} - S_K = \frac{\sqrt{2}}{n}\sum(-b_{i-(K-1)} + b_{i-K}) \qquad (13)$$

giving the scale separation between S_{k-1} and S_k.

The same general procedure may be followed when the equations are of the type

$$S_i + S_j = b_{i+j}\sqrt{2}$$

In this case, however, the table corresponding to (12) is not skew symmetrical, but perfectly symmetrical. That is,

$$\begin{array}{cccccccc} & 1' & 2 & 3 & 4 & & n \\ 1 & — & b_{1+2} & b_{1+3} & b_{1+4} & \cdots & b_{1+n} \\ 2 & b_{1+2} & — & b_{2+3} & b_{2+4} & \cdots & b_{2+n} \\ 3 & b_{1+3} & b_{2+3} & — & b_{3+4} & \cdots & b_{3+n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ n & b_{1+n} & b_{2+n} & b_{3+n} & b_{4+n} & \cdots & — \end{array} \tag{14}$$

This is due to the fact that we are considering sums of scale values rather than differences. Furthermore, instead of having a diagonal of zeros the diagonal elements are indeterminate, and hence left blank.

In forming the difference table of (14) each subsequent column is subtracted from the preceding one, so that each remainder in a given column is of the type $b_{i+(K-1)} - b_{i+K}$. Summing each difference column and multiplying by $\frac{\sqrt{2}}{n-2}$ gives,

$$S_{K-1} - S_K = \frac{\sqrt{2}}{n-2} \sum (b_{i+(K-1)} - b_{i+K}) \tag{15}$$

The set of scale separations obtained from (15) may be averaged with those obtained from (13).

2. *The Least Square Method.*—But while either (13) or (15), or a combination of the two, give scale separations, they give no clue as to the origin of the scale. Since our fundamental problem is the determination of this absolute origin, we must develop a method for combining the sets of equations

$$S_i - S_j = b_{i-j}\sqrt{2} \tag{9}$$

$$S_i + S_j = b_{i+j}\sqrt{2} \tag{11}$$

so that the S's can be determined in terms of their distance from this absolute origin. To determine the scale values in terms of an absolute origin a "least squares" procedure is employed. The normal equations are derived from the observation equations in the conventional manner, but due to the form of the observation equations the normal equations are much easier to derive than with the ordinary (n)-variable problem. The derivation is as follows:

$$\Sigma(\overline{S_i - S_j} - b_{i-j}\sqrt{2})^2 = V = \text{a minimum} \tag{16}$$

$$\Sigma(\overline{S_i + S_j} - b_{i+j}\sqrt{2})^2 = U = \text{a minimum} \tag{17}$$

Considering first equation (16), we have

$$\frac{\partial V}{\partial S_1} = 0, \frac{\partial V}{\partial S_2} = 0, \cdots \frac{\partial V}{\partial S_n} = 0$$

The general form for any equation (K) is,

$$\frac{\partial V}{\partial S_K} = -S_1 - S_2 \cdots + S_K(n-1) \cdots - S_n - \sqrt{2}\left(\sum_{K+1}^{n} b_{ki} - \sum_{1}^{K-1} b_{Kj}\right) = 0 \quad (18)$$

There will be (n) of these equations. First we determine the nature of the matrix of the coefficients (18), leaving the discussion of the formation of the constant terms for a latter paragraph. An examination of (18) shows that the matrix of the coefficients of the parameters or scale values takes the form

$$A_{i-j} = \begin{Vmatrix} n-1 & -1 & -1 \cdots & -1 \\ -1 & n-1 & -1 \cdots & -1 \\ -1 & -1 & n-1 \cdots & -1 \\ \cdots & \cdots & \cdots & \cdots \\ -1 & -1 & -1 \cdots & n-1 \end{Vmatrix} \quad (19)$$

But since the number of -1's in every row of (19) is $(n-1)$, the sum of the elements of every row is zero. According to the theorem in the elementary theory of equations that if the sum of the elements of every row is zero, the determinant of the matrix vanishes, A_{i-j} is zero and there is no unique solution for the constants in equation (18). This necessitates a combination with the normal equations derived from (17). The method of combining the two sets of equations will be discussed later.

Recurring now to the formation of the constant term in (18), it will be noted that the general expression for the K'th constant term is

$$K_K = \sqrt{2}\left(\sum_{K+1}^{n} b_{ki} - \sum_{1}^{K-1} b_{Kj}\right) \quad (20)$$

But we see that (20) is precisely the result obtained if we sum the K'th column in table (12) and multiply by $\sqrt{2}$, where the summations take into account that all the b's above the diagonal have a negative

sign. This procedure enables us to get the respective constant terms directly from the table of b's.

As indicated, however, equations (18) do not give a solution of the parameters, due to the fact that the determinant of the matrix A_{i-j} in (19) vanishes. Now let us consider the normal equations to be derived from (17).

Taking partial derivatives with respect to each of the S's and equating to zero gives (n) equations. These are of the form

$$\frac{\partial U}{\partial S_K} = S_1 + S_2 \cdots + S_K(n-1) \cdots + S_n - \sqrt{2}(\Sigma b_{Ki}) = 0 \quad (21)$$

As in (18) the constant terms of these equations are formed by summing the columns in the table of b_{i+j}'s and multiplying by $\sqrt{2}$.

We see that while the constant terms for (18) are given by

$$K_{i-j} = \sqrt{2}\left(\sum_{K+1}^{n} b_{Kj} - \sum b_{Ki}\right) \quad (20)$$

the constant terms for (21) are given simply by

$$K_{i+j} = \sqrt{2}(\Sigma b_{Kj}) \quad (22)$$

The reason, of course, that we have a difference of two summations in (20) and not in (22) is due to the fact that while in table (12) all the b's above the diagonal are given the negative sign, in table (14) they retain the positive sign. That is, the table is perfectly symmetrical. The method of obtaining the constant terms by summing rows is identical in both cases.

Now let us consider the matrix of the coefficients of (21). An examination of these equations shows that the matrix is given by

$$A_{i+j} = \begin{Vmatrix} n-1 & +1 & +1 \cdots & +1 \\ +1 & n-1 & +1 \cdots & +1 \\ +1 & +1 & n-1 \cdots & +1 \\ \cdot & \cdot & \cdot & \cdot \\ +1 & +1 & +1 \cdots & n-1 \end{Vmatrix} \quad (23)$$

A comparison of (19) and (23) shows that they differ only in that the 1's in (19) are negative, while in (23) they are positive. In both matrices the diagonal terms are the same, viz., $(n-1)$.

Now the determinant of (23) does not vanish, so that equations (21) determine uniquely the values of the S's. However, in an experi-

mental set-up it is usually not feasible to get a matrix as indicated in (23). This is due to the fact that for some situations the proportion of judgments $S_i + S_j > 0$ is close to either zero or unity and hence too unreliable to use. For the same reason the matrix (19) obtained from experimental data differs also from that given. In an experimental set-up many of the elements which are -1 and 1 in (19) and (23) respectively are zero. When this is true the diagonal terms will not be $(n - 1)$, but for any row the diagonal term will be equal to the number of 1's in that row. If it were possible to obtain complete equations of the types (18) and (21) respectively, then the K'th equation of (18) could be added to the K'th equation of (21) giving simply

$$S_K = \frac{\sqrt{2}}{2(n-1)}\left[\left(\sum_{K+1}^{n} b_{K-i} - \sum_{1}^{K-1} b_{K-j}\right) + \sum_{1}^{n} b_{K+i}\right] \tag{24}$$

so that S_K could be obtained directly by adding the summation of the K'th column in (12) to the summation of the K'th column in (17) and multiplying this sum by $\frac{\sqrt{2}}{2(n-1)}$. If we have $\frac{n(n-1)}{2}$ observation equations of the type $S_i - S_j = b_{i-j}\sqrt{2}$ and the same number of the type $S_i + S_j = b_{i+j}\sqrt{2}$ then (24) is the expression which makes $\sum(\overline{S_i \pm S_j} - b_{i\pm j}\sqrt{2})^2$ a minimum where the sign $\pm$ indicates that the summation is through both the $S_i - S_j$ and the $S_i + S_j$ series.

As a matter of fact, in an experimental set-up for equations of the type $S_i + S_i = b_{i+j}\sqrt{2}$ the b's can be obtained only for $S_i + S_j$ when

$$S_i > 0 > S_j$$

or

$$S_i < 0 < S_j \tag{25}$$

The inequalities (25) may not be required providing the ratio S_i/σ_i is comparatively small. If these do not obtain, then (23) takes the form

$$A_{i+j} = \begin{Vmatrix} a_1 & . & 0 & 1 & & . & 1 \\ . & . & . & . & . & . & . \\ 0 & . & a_m & 1 & & . & 1 \\ 1 & . & 1 & a_{m+1} & & . & 0 \\ . & . & . & . & . & . & . \\ 1 & . & 1 & 0 & & . & a_n \end{Vmatrix} \tag{26}$$

where the S's corresponding to all terms up to (m) are negative and those from $(m + 1)$ to (n) are positive, but where the a's are still equal to the sum of the 1's in their respective rows.

Now (26) vanishes, so that if the experimental data result in a determinant of this type there is no solution for the S's in (21).

We next note that when the matrices obtained from the experimental data vary from (19) and (23) respectively this variation is due to the fact that certain of the b's in tables (12) and (14) are missing. But knowing which of these b's are missing the matrices can always be set up directly from the tables.

But to determine the S's uniquely, equations (18) and (21) are added together yielding a matrix which is the sum of (19) and (23).

The normal equations may be set up directly from tables (12) and (14) as follows:

To get the diagonal element for any row, count the total number of entries for that row in the corresponding rows of both tables. This number will be the diagonal element. To get any other element:

(1) If its corresponding row and column has an entry for both (12) and (14) the element is zero.

(2) If its corresponding row and column lacks an entry in both (12) and (14) the element is zero.

(3) If its corresponding row and column has an entry in only table (12), the element is -1.

(4) If its corresponding row and column has an entry in only table (14), the element is $+1$.

In a numerical set up, only those terms above the diagonal need be written down.

The constant term for any row is obtained by adding the sum of the elements in the corresponding column in table (12) to the sum of the elements in the corresponding column of table (14) and multiplying this total by $\sqrt{2}$.

The equations may be solved most conveniently by the Doolittle method (4). The solution is considerably simplified since all of the coefficients except the diagonal terms are either plus or minus 1, or zero.

3. *Correspondence of Calculated with Observed Values.*—To find the degree of accuracy with which the solutions fit the experimental data, we determine the average squared deviation in the equations

$$(\overline{S_i - S_j} - b_{i-j}\sqrt{2})^2 = d^2{}_{i-j}$$

and

$$(S_i + S_j - b_{i-j}\sqrt{2})^2 = d^2_{i+j}$$

where the values obtained for the S's are substituted in the equations.

The value d_{i-j} is the discrepancy obtained by subtracting the observed difference between the scale values of two stimulus situations from the calculated difference. The value d_{i+j} has the same meaning for the sum of two scale values. The square root of the average of all of these deviations is called the standard error of estimate and is written

$$\sigma_{est.} = \sqrt{\frac{\Sigma d^2}{n}}$$

where (n) is the total number of both sum and difference equations. Of course, $\sigma_{est.}$ may be reduced to $PE_{est.}$ by the relation

$$PE_{est.} = 0.6745\sigma_{est.}$$

Now in any regression equation of the general form

$$Y = B_1x_1 + B_2x_2 + \cdots + B_nx_n$$

the $PE_{est.}$ means that approximately fifty per cent of the experimental values will not deviate from the corresponding calculated values by more than the numerical value of the $PE_{est.}$. It may be interpreted as a measure of the error in the prediction or regression equation.

But in our problem the regression equation takes the simpler form

$$S_i - S_j = b_{i-j}\sqrt{2}$$

or

$$S_i + S_j = b_{i+j}\sqrt{2}$$

Here, then, the PE is a measure of the error in a sum or difference of two values. But the error of a sum or difference of two values is equal to the square root of the sum of the squares of their respective errors.

Hence, if $PE_{est.} \equiv PE_{ij}$ is the error in $S_i - S_j$ or $S_i + S_j$ and PE is the error in any given scale value, then

$$PE_{ij} = \sqrt{(PE_i)^2 + (PE_j)^2} \qquad (27)$$

If we assume that the errors in all the scale values are equal we obtain from (27)

$$PE_i = \frac{PE_{ij}}{\sqrt{2}} \qquad (28)$$

To get PE_{ij} we need not actually find the residual for each equation and then square and sum. From the method of fitting it follows that for any "least square" solution of the type

$$\Sigma(\overline{B_1x_1 + B_2x_2 + \cdots B_nx_n} - x_0)^2 = \text{a minimum},$$

that the standard error of prediction (5) is given by,

$$\sigma_{est.} = \left[\frac{\Sigma x_0^2 - (B_1x_0x_1 + B_2x_0x_2 + \cdots B_nx_0x_n)}{n}\right]^{\frac{1}{2}} \tag{29}$$

It will be noted that Σx_0^2 is the sum of the squares of the dependent variable. In our problem this is

$$\Sigma x_0^2 = 2\Sigma b^2{}_{i \pm j} \tag{30}$$

where the sign $\pm$ indicates summation through both the difference and sum series. Furthermore,

$$B_1 = S_1,\ B_2 = S_2,\ \cdots B_n = S_n \tag{31}$$

and

$$\Sigma x_0x_1 = K_1,\ \Sigma x_0x_2 = K_2,\ \cdots \Sigma x_0x_n = K_n \tag{32}$$

Substituting values (30), (31), and (32) in (29), we have

$$\sigma_{est.} = \left[\frac{2\Sigma b^2{}_{ij} - \Sigma S_iK_i}{n}\right]^{\frac{1}{2}} \tag{33}$$

or from (28)

$$PE_i = 0.6745\sqrt{\frac{2\Sigma b^2{}_{ij} - \Sigma S_iK_i}{2n}} \tag{34}$$

Equation (32) is the expression for the PE of a scale value.

To obtain the value $\Sigma b^2{}_{ij}$, tables are made up from tables (12) and (14) by squaring the values in these tables. Only the sections above the diagonals are included in the tables of squares, as those below have the same value when squared. The sums of the rows of both tables are added together to get $\Sigma b^2{}_{ij}$.

The value ΣS_iK_i is readily obtained by multiplying each constant term by the corresponding scale value and summing.

IV. EXPERIMENTAL

1. *Method of Obtaining Values for the Difference and Sum Equations.* The equations to be kept in mind in an experimental set-up are

$$P_{i-j} = \int_{-\infty}^{x_{i-j}-(S_i-S_j)} f(x_{i-j})dx_{i-j} \tag{7}$$

with its solution for $S_i - S_j$,

$$S_i - S_j = b_{i-j}\sqrt{2} \tag{9}$$

and

$$P_{i+j} = \int_{-\infty}^{x_{i+j}-(S_i+S_j)} f(x_{i+j})dx_{i+j} \tag{10}$$

with its solution

$$S_i + S_j = b_i\sqrt{2}_{+j} \tag{11}$$

The method for determining P in (7) is as follows:

A series of stimulus situations is given ranging in affective value from distinctly positive to distinctly negative. Denoting the situa-

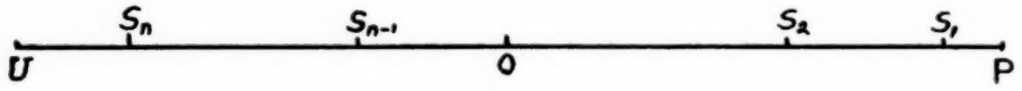


FIG. 7.

tions by S_1, S_2 to S_n, we may represent them diagrammatically on a scale as in Fig. 7:

Now when we say that situation 1 is more pleasant than situation 2 for a given individual, we mean that he prefers 1 to 2. By definition, 1 has for him a higher affective value than 2. Hence, to find whether situation i is greater in affect than situation j, statements of the situations are assembled according to the method of paired comparisons on mimeographed sheets with phrasing as follows:

( ) 1–2.	I would rather have S_1 than S_2.
( ) 1–3.	I would rather have S_1 than S_3.
. . . . .	. . . . .
( ) 1–n.	I would rather have S_1 than S_n.
( ) 2–3.	I would rather have S_2 than S_3.
. . . . .	. . . . .
( ) 2–n.	I would rather have S_2 than S_n.
. . . . .	. . . . .
( ) (n–1)–n.	I would rather have S_{n-1} than S_n.

These are presented to a group of individuals. If the individual considers the statement true for himself, he marks (+) in the parenthesis. If not, he marks (−), just as in a True-False test. He is

instructed to leave no blanks. For each statement of paired situations, then, there is a certain proportion of individuals who marked it (+). This is the value of P in (7) and from which the value b_{i-j} in (9) is obtained from a table of the normal probability integral. These b_{i-j} values are the ones entered in table (12).

When we say that situation 1 is more pleasant for a given individual than (n) is unpleasant, we mean that if he were permitted to enjoy 1, he would be willing to submit to (n). By definition, the cumulative affect of the two situations is, for him, positive. Hence, to find how much the cumulative affect of situations i and j differ from zero, the situations are assembled on a second set of mimeographed sheets with phrasing as follows:

( ) 1–2.	If I could have S_1 I would be willing to take S_2.
( ) 1–3.	If I could have S_1 I would be willing to take S_3.
. . . . .	. . . . .
( ) 1–n.	If I could have S_1 I would be willing to take S_n.
( ) 2–3.	If I could have S_2 I would be willing to take S_3.
. . . . .	. . . . .
( ) 2–n.	If I could have S_2 I would be willing to take S_n.
. . . . .	. . . . .
( ) (n–1)–n.	If I could have S_{n-1} I would be willing to take S_n.

These are presented to the same group of individuals as was the first series. Again for each statement of paired situations, the proportion of individuals who marked it (+) is obtained. This percentage is the value of P_{i+j} in (10) and from which the value b_{i+j} in (11) is obtained from a normal probability table. These b_{i+j} values are the ones entered in table (14).

On the mimeographed forms for the first series, however, pairs of situations which are obviously far apart on the scale may be omitted, for they would result in proportions either too high or low to be reliable. For the same reason in the second series, situations which are obviously close together, and clearly some distance from the point of zero desirability or affect, are also omitted from the series.

In the experimental series the situations were obtained from a number of the individuals who judged the pairs, as follows:

The subjects were one hundred sixty students in a teachers college. The following instructions were given:

> Describe in one sentence a situation which you consider extremely unpleasant, one which you consider extremely pleasant, one which you consider as being

entirely neutral, another which is moderately unpleasant, and still another which is moderately pleasant.

Thus each individual submitted five situations. In all, about two hundred situations were obtained. The experimenter then selected twelve out of these two hundred according to two criteria:

(1) Subjectively they appeared to be relatively uniformly extended through a scale from extreme unpleasantness to extreme pleasantness.

(2) When possible, situations were selected which were most frequently described.

The object of (2) was, of course, to select situations which were relevant to the everyday experience of the group which was to judge them. They were then edited so as to fit as logically and grammatically as possible into the form,

I would rather have S_i than S_j,

and

I would be willing to take S_i if I could have S_j, respectively.

The situations which were put in the mimeographed forms were phrased as follows:

1. Receive a kiss from the one I love.
2. Dance with a good dancer.
3. Go to a good musical comedy.
4. See a beautiful art exhibit.
5. Receive a smile from a friend of the opposite sex.
6. Watch a strange man try to choose between a red and a green necktie.
7. Lose several hairs from my head.
8. Hear a beginner practise his violin lesson.
9. Discover I had forgotten my purse after having purchased an article.
10. Have my clothes splashed by a passing vehicle when going to a party.
11. See my sweetheart come into the show with another date.
12. Spend a night in jail.

2. *Data and Calculations.*—Table I gives the (b) values for both the sum and the difference series. Table II shows the normal equations set up from Table I in the manner previously outlined, and ready for solution. The scale values, as determined from a solution of these equations is given in Table III.

To find the PE of a scale value, we employ the formula

$$\text{PE} = 0.6745\sqrt{\frac{2\Sigma b^2_{ij} - \Sigma S_i K_2}{2n}} \qquad (32)$$

To get Σb^2_{ij} another table is constructed from Table I by squaring only those b's above the diagonal. Multiplying the sum of all the b^2's by 2, we get,

$$2\Sigma b^2_{ij} = 150.6480$$

The value $\Sigma S_i K_i$ is obtained by multiplying column K of Table II by the column of scale values in Table III and summing the products. We find that

$$\Sigma S_i K_i = 145.5676$$

The number of observation equations is

$$N = 56.$$

Substituting in the formula

$$\text{PE} = 0.6745\sqrt{\frac{150.6480 - 145.5676}{112}} = 0.1442$$

The PE of a scale value, then, is 0.1442.

TABLE I.—SIGMA VALUES OF PROPORTIONS

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	Σ
Series 2													
1	.....	.....	.....	.....	.....	.....	.....	.....	− .34	− .78	− .99	− 1.51	−3.62
2	.....	.....	.....	.....	.....	.....	.....	.47	− .61	−1.19	− .98	− 2.01	−4.32
3	.....	.....	.....	.....	.....	.....	.....	1.39	− .21	− .94	− .79	− 1.96	−2.51
4	.....	.....	.....	.....	.....	.....	− .45	.87	− .61	−1.10	−1.08	.....	−1.47
5	.....	.....	.....	.....	.....	.....	− .45	− .05	−1.23	−1.55	−1.67	.....	−4.95
6	.....	.....	.....	.....	.....	.....	− .91	− .70	−1.77	.....	.....	.....	−3.38
7	.....	.....	.....	.45	− .45	− .91							
8	.....	.47	1.39	.87	− .05	− .70							
9	− .34	− .61	− .21	− .61	−1.23	−1.77							
10	− .78	−1.19	− .94	−1.10	−1.55								
11	− .99	− .98	− .79	−1.08	−1.67								
12	−1.51	−2.01	−1.96										
Series 1													
1	.....	− .58	− .04	− .39									
2	.58	.....	.38	.06	− .66								
3	.04	− .38	.....	− .50	−1.26	−1.57							
4	.39	− .06	.50	.....	− .85	−1.38	−1.85						
5	.....	.66	1.26	.85	.....	−1.07	−1.85						
6	.....	.....	1.57	1.38	1.07	.....	−1.63	−1.28	−1.70				
7	.....	.....	.....	1.85	1.85	1.63	.....	.44	− .63	−1.17			
8	.....	.....	.....	.....	.....	1.28	− .44	.....	−1.35	−1.90	−1.54		
9	.....	.....	.....	.....	.....	1.70	.63	1.35	.....	− .58	− .77	− 2.00	
10	.....	.....	.....	.....	.....	.....	1.17	1.90	.58	.....	− .42	− 1.63	
11	.....	.....	.....	.....	.....	.....	.....	1.54	.77	.42	.....	1.11	
12	.....	.....	.....	.....	.....	.....	.....	.....	2.00	1.63	1.11		
Total.....	−2.61	−4.68	1.16	1.78	−4.80	−2.79	−4.88	5.93	−5.10	−7.16	−7.13	−10.22	

TABLE II

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	K	Σ
1	7.0000	−1.000	−1.000	−1.000	.....	.....	.....	.....	1.000	1.000	1.000	1.000	3.6911	11.691
2	.....	9.000	−1.000	−1.000	−1.000	.....	.....	1.000	1.000	1.000	1.000	1.000	6.6185	16.618
3	.....	.....	10.000	−1.000	−1.000	−1.000	.....	1.000	1.000	1.000	1.000	1.000	−1.6405	8.359
4	.....	.....	.....	11.000	−1.000	−1.000	.....	1.000	1.000	1.000	1.000	.....	−2.5173	7.483
5	.....	.....	.....	.....	10.000	−1.000	.....	1.000	1.000	1.000	1.000	.....	6.7882	16.788
6	.....	.....	.....	.....	.....	9.000	.....	.....	.....	.....	.....	.....	3.9456	9.946
7	.....	.....	.....	.....	.....	.....	9.000	−1.000	−1.000	−1.000	.....	.....	6.9013	12.901
8	.....	.....	.....	.....	.....	.....	0	10.000	−1.000	−1.000	−1.000	.....	−8.3862	1.6138
9	.....	.....	.....	.....	.....	.....	.....	.....	12.000	−1.000	−1.000	−1.000	7.2124	19.212
10	.....	.....	.....	.....	.....	.....	.....	.....	.....	10.000	−1.000	−1.000	10.1257	20.126
11	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	9.000	−1.000	10.0832	20.083
12	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	6.000	14.4531	20.453

TABLE III

SITUATION	ABSOLUTE SCALE VALUES
3	2.49
1	2.48
4	1.93
2	1.90
5	1.03
6	0.17
8	−1.01
7	−1.55
9	−2.64
10	−3.39
11	−3.57
12	−5.15

V. SUMMARY

1. The scaling method developed by Thurstone is described, and the assumptions underlying the method are presented.

2. The equations required in solving for the absolute affective values of a series of stimulus situations are derived analytically. The derivations include the assumptions involved in the method of scaling.

3. Two methods for determining scale values are presented:

(*a*) The "average difference" method yields only scale separations and necessitates the arbitrary selection of an origin.

(*b*) The "least square" method gives a solution of the affective values of stimulus situations in terms of their distances from an absolute zero.

4. The normal equations for the least square solution are derived and their typical form discussed.

5. Instructions are outlined for writing the normal equations directly from the table of sigma values of the experimental proportions.

6. Statistical procedure is presented for determining how accurately the calculated scale values fit the experimental data.

7. An experimental method is devised for obtaining data in the form required for the scaling equations.

8. To demonstrate the scaling method a model experiment was conducted; the data resulting from the experiment are analyzed according to the method outlined and the results of the calculations are presented.

VI. REFERENCES

1. Thurstone, L. L.: Psychophysical Analysis, *American Journal of Psychology*, 1927, Vol. XXXVIII, pp. 368–389.
2. Hanus, P. H.: "Elementary Theory of Determinants."
3. Dickson, L. E.: "First Course in the Theory of Equations," p. 116.
4. Mills, Frederick: "Statistical Methods," 577 f.
5. Mills, Frederick: "Statistical Methods," pp. 570.
6. Granville, N. A.: "Differential and Integral Calculus," pp. 242.

VISUAL IMAGERY AND "OBSERVATION"

HENRY BOWERS

Ottawa Normal School

It has been pointed out[1] that certain stimuli evoking visual, auditory, or kinaesthetic images fall into a relatively constant order according to the mean ratings assigned by different groups to the aroused images. This suggested the possibility that the mean ratings assigned by groups to images evoked by the various items in a picture would show a similar relatively constant order of precedence. If this were so it would be interesting to compare this order with that into which the items fall when considered from the standpoints (*a*) frequency of mention in "reports," (*b*) mean ranks in "reports."[2]

Problems.—1. To determine if there is a relatively constant order of precedence for the items in a picture on the basis of clarity of the visual imagery that can be aroused by these items.

2. If this relatively constant order of precedence exists to compare it with those previously established for frequency of mention of items and for mean order of mention in "reports."

3. If the order of precedence mentioned in (1) above exists, and if significant correlations are found between this order and those referred to in (2) above to determine if the relationship is a causal one.

Procedure—Problem (1).—The subjects used were first and second year high school pupils and the test material was a coloured picture called "A Mountain River," one in flat tones of a Labrador fiord, and Test 6 from the writer's "Transfer Values of Secondary School Science."[3] The latter test consists of thirteen exercises contained in a booklet. On one side of each page is a geometrical or nonsense figure; the back of the page contains a figure which could be made identical with the preceding one by the addition of certain details. It is the

---

[1] Bowers, H.: Constancy of Imaginal Content. *Journal of Educational Psychology*, Vol. XX, 1929, pp. 295–298.

——: Memory and Mental Imagery: An Experimental Study. *Brit. J. Psychol.*, Vol. XXI, 1931, pp. 271–282.

——: Studies in Visual Imagery. *American Journal of Psychology*, Vol. XLIII, 1931, pp. 216–229.

[2] Bowers, H.: Order and Frequency of Observations Recorded by a Group. *Journal of Educational Psychology*, Vol. XXIII, 1932, pp. 221–225.

[3] Bowers, H.: "Transfer Values of Secondary School Science." The University of Toronto Press, Toronto, 1927.

subject's task to supply the details after examination of the first figure. Ten minutes is allowed for the thirteen exercises. Scores on the test made by six hundred twenty-one third- and fourth-year high school pupils are distributed in a close approximation to the normal probability curve. The mean reliability coefficient for a single grade is .50 ± .07. In scoring the test one point is allowed for each correctly completed exercise. That this procedure is legitimate may be inferred from the correlation coefficient .95 ± .01 found to exist between "sum of correct elements" and "sum of scores for correct elements." The latter scores were obtained by weighting each element for difficulty in the usual manner.

"Reports" for each of the two pictures were obtained as described in "Order and Frequency of Observations Recorded by a Group." It may be mentioned that two minutes were allowed for examination of the picture and at the end of this period the subjects were required to write everything that they could recall about it. For present purposes only substantives were scored; attributives were ignored. After these statements were scored a list was prepared of the recorded items. Each subject was supplied with a copy of this list and a copy of the picture. He was requested to take each item in turn, find it in the picture, and examine it closely. Then he was to close his eyes and evoke the image. This image was rated on a four-step scale. Values 0, 1, 2, and 3 were arbitrarily assigned to the four stages. Each subject had had previous experience in rating visual images.

After scoring Test 6 the subjects were requested to rate the images aroused by each of the completed figures.

The mean rating for each item in both pictures and for each element in Test 6 was found for random halves of the groups tested. Reliability coefficients for these mean ratings are given in Table I.

TABLE I.—RELIABILITY COEFFICIENTS FOR MEAN RATINGS OF IMAGES

Test material	No. of subjects in each group	No. of items	
"Mountain River"	24 and 24	19	.84 ± .04
"Labrador Fiord"	24 and 24	16	.69 ± .09
Test 6	24 and 24	13	.91 ± .03

These coefficients extend to the constituent elements of pictures the property of relative constancy of imaginal content shown to exist for other material. It is obvious that the mean estimate for the whole

group of forty-eight is much more stable than those for the two subgroups of twenty-four.

It is interesting to compare this property of the elements composing a picture with E. R. Jaensch's statements concerning the suitability or unsuitability of certain pictures for the discovery and investigation of eidetic imagery. He says[1] . . . "the appearance of the conceptual factor is not so dependent on fixation or on factors that influence the sense organ. It depends rather on the interest taken in the picture and on the involuntary attention aroused by the picture . . . Certain types of objects are preferred and sometimes give rise to clearer *EI*s[2] by themselves quite independently of the purely optical qualities and particularly of the sensory intensity of the picture. Animal pictures pick out a higher percentage of *EI*s from children." In another place he criticizes Schwab's[3] use of a test picture as "it fails to conform to almost every condition for a satisfactory test-object. It lacks (1) a sensible coherence; (2) it contains a large number of similar or approximately similar elements; (3) there are a number of regions of the same colour and shape sometimes symmetrically arranged."

Problem (2).—The groups which provided "reports" on both pictures were divided into halves by random selection. For each half the mean frequency and mean rank were calculated for items appearing in these "reports." Reliability coefficients for these mean frequencies and mean ranks are reported elsewhere ("Order and Frequency of Observations Recorded by a Group"). The mean r for frequency was .85, that for rank, .70. Table II shows the relationship found between imagery rating, mean frequency, and mean rank.

TABLE II.—CORRELATION BETWEEN MEAN RATING OF ITEM FOR IMAGERY AND MEAN FREQUENCY IN REPORT. CORRELATION BETWEEN MEAN RATING OF ITEM FOR IMAGERY AND MEAN RANK IN REPORT

Picture	No. of subjects	No. of items	r (imagery and frequency)	r (imagery and rank)
"Mountain River".....	48	19	$+.59 \pm .10$	$-.75 \pm .07$
"Labrador Fiord".....	28	16	$+.36 \pm .15$	
Labrador Fiord.....	28	13	.....	$-.48 \pm .16$

[1] Jaensch, E. R.: "Eidetic Imagery." Kegan, Paul, Trench, Trubner & Co., 1930.

[2] Eidetic images.

[3] Schwab, G.: Vorläufige Mitteilung über Untersuchungen zum Wesen der subjektiven Anschauungsbilder. *Psychol. Forschung*, Vol. V, 1924.

The coefficient of correlation between the mean score for each of the thirteen elements in Test 6 and the mean imagery rating for each of these exercises was $+.62 \pm .12$ (forty subjects). Another determination gave $r = +.57 \pm .13$ (fifty subjects). Each of these coefficients was obtained from a single grade. It is estimated that correction for attenuation should increase all of these *r*s by approximately one-fifth.

It is apparent that there is an association between mean frequency of mention in "reports" and mean rating for aroused imagery. There is also a tendency for a group to record those items in "reports" in an order approximating to that in which they were rated for their effectiveness in evoking visual images. It would also appear that there is a tendency for those exercises in Test 6 which lent themselves most readily to reproduction by visual imagery to be solved most easily.

Problem (3).—Although significant correlation exists between the facility with which the component items of a picture are reproduced in visual images and the frequency and rank of these items in "reports" it cannot be concluded that clarity of imagery determines even partially this rank and frequency. Further data are necessary.

The writer's test of visual imagery ("Studies in Visual Imagery") was given to thirty-seven subjects from whom "reports" on "Mountain River" had been obtained. Total scores on the visual imagery test and total scores on the "reports" gave $r = +.22 \pm .11$. The correction for attenuation would raise this coefficient to approximately .25. (r_{11} for the test in visual imagery is .96 for a single grade, that for the "report" is certainly not above .8). Two values were found of the coefficient of correlation between scores on Test 6 and scores on the test for visual imagery. They are $+.15 \pm .10$ ($N = 40$) and $+.05 \pm .12$ ($N = 32$).

No correspondence can be detected between the ability of a subject to evoke clear visual images of pictorial representations and the ability to remember these and record them. In view of the established relationship between frequency and rank ("Order and Frequency of Observations Recorded by a Group") it may be assumed that the relationship between imagery and order is not causal.

Remarks on Method.—1. An apparent limitation of our technique is that although only two minutes was allowed for examination of the picture before a "report" was required, as much time as desired was given when rating the clarity of mental imagery aroused by each item. In defence of this is the *r* of $+.86 \pm .06$ obtained between total scores

on "reports" in which a two-minute observation period was allowed and those secured when using a five-minute period. Eighty-three days separated the two sets of reports.

2. It is unfortunate that more items were not recorded. This is due of course to the nature of the pictures and also to the limitations of memory. The calculation of mean rank in the case of several items was impossible in view of the extreme unreliability involved in the use of very few cases. These items were discarded.

3. A certain degree of subjectivity is involved in scoring the reports. However, in view of the fairly consistent coefficients obtained it would seem that in spite of the difficulties referred to above, our conservative conclusions have adequate basis.

The matter of subjectivity in the test for visual imagery has been discussed elsewhere.[1]

Conclusions

1. The constituent elements of a picture exhibit that relative constancy of imaginal content found to exist for other visual, auditory, and kinaesthetic stimuli.

2. A small but significant correlation exists between mean ratings of images aroused by items in a picture and frequency of mention of these items in "reports" of the picture.

3. A rather more pronounced association exists between clarity of image and mean rank of item.

4. There is evidence that the relationships referred to in (2) and (3) are not of a causal nature.

Evidence supporting conclusions (1), (2) and (4) was obtained from a test involving the completion of geometrical and nonsense figures.

---

[1] Bowers, H.: Studies in Visual Imagery. *Am. J. Psychol.*, Vol. XLIII, 1931, pp. 271–282.

"STUDENT'S" METHOD FOR MEASURING THE SIGNIFICANCE OF A DIFFERENCE BETWEEN MATCHED GROUPS

MORDECAI EZEKIEL

Federal Farm Board

In a recent article,[1] Lindquist shows that when the effectiveness of two methods of teaching are compared by the use of "matched groups," the correct probable error of the difference between the mean of two groups may really be very much smaller than the figure computed by the usual formula for the error in the difference of means. He further explains a new formula, developed by Wilks,[2] the "standard error of the mean of matched samples," which serves as a means of estimating the increased significance of the results obtained by the "matching" process.

As is not unusual in the progress of science, the development of this method by Wilks and Lindquist represents a rediscovery, by workers in one field, of what has long been known by workers in another. Over twenty years ago, an eminent English statistician "Student" (he writes under this pseudonym) considered the same type of problem which Lindquist attacks, and arrived at parallel conclusions.[3] As it happens, however "Student" developed a method of measuring the accuracy of the mean difference between the two matched groups which requires much less work to compute than does the Wilks-Lindquist formula. With the thought that psychologists and educators might be able to save much energy by the use of this simpler method, this article is prepared to call their attention to its existence.

The method proposed by Lindquist will be illustrated by a set of numerical data, and then "Student's" method will be shown by applying it to the same data.

---

[1] Lindquist, E. F.: The Significance of a Difference between "Matched" Groups. *Journal of Educational Psychology*, Vol. XXII, 1931, pp. 197–204.

[2] Wilks, Samuel S.: The Standard Error of the Means of "Matched" Samples. *Journal of Educational Psychology*, Vol. XXII, 1931, pp. 205–208.

[3] Fisher, R. A.: "Statistical Methods for Research Workers." Second edition, 1928, pp. 104–122.

"Student": The Probable Error of a Mean. *Biometrika*, Vol. VI, 1908, pp. 1–25.

We will use the same example that Lindquist considers, an experiment to determine the relative effectiveness of two methods of teaching subtraction, with one group of children taught by the "additive" method, and another by the "take-away" method. After a period of instruction, a final measure of ability in subtraction would be obtained for all the pupils in both groups. To reduce random errors in the experiment, the two groups would be "matched" by trying to have the average initial ability of the pupils the same in each group. "If the intelligence test scores had been used, the two groups would have been so selected as to have identical distributions of intelligence test scores; in other words, they would have been 'matched' on the basis of intelligence."[1]

The following set of hypothetical data illustrate the results that might be secured by an experiment using the "matching" method. (Only ten cases are shown in each sample; in practice, larger groups would ordinarily be required for dependable results.)

TABLE I.—HYPOTHETICAL DATA ILLUSTRATING RESULTS SECURED BY "MATCHING" PROCESS

Group A		Group B	
Initial IQ	Final rating in subtraction	Initial IQ	Final rating in subtraction
80	58	75	65
90	63	88	67
95	83	98	85
98	78	99	83
103	80	103	86
105	88	107	88
110	85	108	90
122	86	119	89
135	91	138	96
160	94	155	98
Averages....109.8	80.6	109.0	84.7

The data for the children in both groups are arranged in order of increasing IQ, from lowest to highest. In addition to the average IQ being almost exactly the same for both groups, the children have been so selected that the distribution is nearly identical in the two groups,

[1] Lindquist: *Loc. cit.*

for each child in the first group there being a corresponding child in the second group, with about the same IQ.

The average rating on subtraction for Group A is 80.6, and for Group B is 84.7. Is the difference of 4.1 points significant evidence that the method used in teaching Group B is superior to the method used in teaching Group A?

By the usual method, the probable error of the difference in the averages between two groups is given by the formula:[1]

$$PE_{1-2} = .6745\sqrt{\frac{\sigma_1^2}{n_1 - 1} + \frac{\sigma_2^2}{n_2 - 1}}$$

(this differs from the formula frequently given by substituting $n - 1$ for n.)

When the number of cases in both groups is the same, (as in this case) the formula becomes

$$PE_{1-2} = \sqrt{PE_1^2 + PE_2^2}$$

$$= .6745\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n_1 - 1}}$$

For the example shown, the standard deviation of the subtraction scores (σ_S) is 11.07 for Group A, and 10.26 for Group B. Substituting these values in the last equation, the probable error of the difference between the two averages is found to be

$$PE_{1-2} = .6745\sqrt{\frac{(11.07)^2 + (10.26)^2}{9}}$$

$$= 3.39$$

Using this method, the difference between the scores of the two groups would be stated to be 4.1 ± 3.39, which would be taken to mean that the difference observed could readily have occurred by chance and that therefore there was no significant difference between the results secured by the two methods of teaching.

The Wilks-Lindquist formula recognizes the correlation between the criterion used in "matching" the groups, and the final scores. Using I for IQ, and S for score, (and substituting $n - 1$ for n) the formula is

$$\sigma_{M'} = \frac{\sigma\sqrt{1 - r^2_{IS}}}{\sqrt{n - 1}}$$

where $\sigma_{M'}$ represents the "standard error of means of matched groups."

[1] Fisher, R. A.: "Statistical Methods for Research Workers." P. 109. (This is simplified from the form as Fisher gives it.)

For the standard error of the difference between two matched groups, the formula is

$$\sigma_{1-2} = \sqrt{\left(\frac{\sigma_1^2}{n_1 - 1} + \frac{\sigma_2^2}{n_1 - 2}\right)(1 - r^2)}$$

For the twenty cases in Groups A and B the correlation between IQ and subtraction score is 0.813. Substituting this value in the last equation, the probable error of the difference is:

$$PE_{1-2} = .6745\sqrt{\left[\frac{(11.07^2)}{9} + \frac{(10.26)^2}{9}\right][1 - (0.813)^2]}$$

$$= 1.98$$

Hence by the Wilks-Lindquist method the difference between the scores of the two groups would be 4.1 ± 2.0. While a difference as large as this might still occur by chance, this would indicate a good probability that the method for Group B was superior to that for group A, whereas the usual method indicated that there was a considerable probability that the difference was due solely to chance.

"Student's" method is based upon an entirely different approach. It compares the scores for each child in Group A with the score of the corresponding child in Group B, and asks whether the average of these differences in score is significant when contrasted with the *variation in the individual differences.* It is illustrated in Table II.

TABLE II.—COMPUTATION OF DIFFERENCE IN RATING IN SUBTRACTION FOR "MATCHED" CHILDREN

IQ	Final rating in subtraction		Difference in rating ($B - A$)
	Child in Group A	Child in Group B	
75 and 80	58	65	7
88 and 90	63	67	4
95 and 98	83	85	2
98 and 99	78	83	5
103	80	86	6
105 and 107	88	88	0
108 and 110	85	90	5
119 and 122	86	89	3
135 and 138	91	96	5
155 and 160	94	98	4
Means.....	80.6	84.7	4.1

All that remains to be done is to calculate the standard deviation of the individual differences, $(B - A)$. This comes out 1.90. The probable error of the differences may then be calculated by modifying the formula for the probable error of a mean:

$$PE_{B-A} = .6745\sqrt{\frac{\sigma^2_{B-A}}{n-1}}$$

Which in this case gives the value

$$PE_{B-A} = .6745\sqrt{\frac{(1.90)^2}{9}}$$

$$= .43$$

So using "Student's" method, the difference in average score between the two matched groups would be stated to be 4.1 ± 0.43, indicating that method B is superior to method A, and that it is almost a certainty that the difference could not have occurred as a result of chance fluctuations, *even with samples as small as ten.*[1]

"Student's" method can be applied to matched samples only when they have been so matched that for each case in one sample there is a corresponding case in the other sample, with practically identical characteristics insofar as the "matching" criterion is concerned. Thus in the arithmetic example shown, the two IQ's were almost identical for each pair of corresponding children. If the general distribution of the criterion was the same in each sample, but the case-for-case correspondence did not hold, "Student's" method could not logically be applied.

It is evident that where it can be applied, "Student's" method requires much less computation than does the Wilks-Lindquist method. The latter method requires the computation of two standard deviations and one Pearsonian coefficient of correlation, whereas the former requires the computation of the differences shown in Table II, and of their standard deviation. Furthermore, "Student's" method gives a more exact measure of the reliability of the differences. While the Wilks-Lindquist formula gives a more accurate estimate of the reliability than does the ordinary formula, it still much under-estimates the significance of the results, as judged by "Student's" method.

[1] According to the tables for small samples, a departure as large as this (with the average over six times its standard error) would occur by chance less than once out of one hundred samples. See Fisher, R. A.: "Statistical Methods for Research Workers," p. 139; Mordecai Ezekiel, "Methods of Correlation Analysis," Table A, 1930, p. 20.

STANDARDIZATION OF THE STENQUIST MECHANICAL ASSEMBLYING TEST SERIES III

EDNA WILLIS McELWEE

In addition to standardizing the Stenquist Mechanical Assemblying Test Series III, the purpose of this study was three-fold:

1. To correlate mechanical ability and intelligence.
2. To compare the mechanical performance of dull and bright boys of the same chronological age.
3. To compare the mechanical performance of older and younger boys of the same mental age.

The Stenquist Mechanical Assemblying Test Series III was designed for younger boys. It consists of a plain bolt and nut, a bolt and wing nut, a plain hinge, a key and ring, a turn buckle, a drawer pull, a trunk caster, a plain push button, a belt and buckle and a nail clip.

The test was given to four hundred boys six, seven, eight and nine years of age. There were one hundred boys in each age group. They were enrolled in Grades IA to VB inclusive. As practically every boy of these ages in the school was included in the study, it may be considered a normal group.

The boys were examined in groups of three, because that number of test boxes was available. A maximum time of twenty minutes was allowed to complete the test, but only a few boys spent that long. Many of them reached their limit in ten minutes. No time bonus was given in scoring as is done in Series I and II.

Every boy was able to do something, the lowest score being eighteen. None of the boys made a perfect score (one hundred twenty), the highest being one hundred five. Two boys received full credit and eighteen others partial credit for the nail clip, the most difficult article to be assembled. Therefore, the writer feels that the mechanical ability of all the boys was fully measured.

In Table I the scores made by each age group are tabulated in deciles, thus standardizing the Stenquist Mechanical Assemblying Test Series III and presenting norms for scoring. Although some of the six year old boys make a higher score than some of the nine year old boys, there is a steady gain in mechanical performance as chronological age increases.

TABLE I.—NORMS FOR THE STENQUIST MECHANICAL ASSEMBLYING TEST SERIES III BASED ON FOUR HUNDRED CASES

Decile	Year 6	Year 7	Year 8	Year 9
10	66–87	72–95	81–101	86–105
9	56–65	65–71	72–80	81–85
8	51–55	62–64	67–71	74–80
7	47–50	57–61	63–66	70–73
6	45–46	54–56	60–62	66–69
5	41–44	50–53	56–59	62–65
4	37–40	45–49	50–55	57–61
3	34–36	40–44	47–49	54–56
2	28–33	36–39	40–46	50–53
1	18–27	22–35	31–39	36–49

In connection with another study at the same school, group tests of intelligence had been given to a number of the classes. These IQ's were available for the boys seven, eight, and nine years of age.

TABLE II.—IQ DISTRIBUTION FOR THE BOYS SEVEN, EIGHT AND NINE YEARS OF AGE

Range of IQ	Number of boys		
	Year 7	Year 8	Year 9
70–79	1	19	20
80–89	22	38	28
90–99	32	17	16
100–109	23	14	16
110–119	13	7	7
120–129	8	4	6
130–139	0	0	6
140–149	1	1	1

In Table II the IQ distribution of these three age groups is presented. By arbitrarily regarding the IQ range 70–79 as Borderline, 80–89 Dull, 90–99 Normal, 100–119 Bright and 120–149 Superior, the three hundred boys are grouped as follows: forty Borderline, eighty-eight Dull, sixty-five Normal, eighty Bright and twenty-seven Superior. This grouping suggests a normal curve.

Using the Product Moment formula, correlations were worked out between mechanical ability and intelligence for each of the three age groups.

Seven year group	$r = .102$	PE $\pm$ .067
Eight year group	$r = -.206$	PE $\pm$ .064
Nine year group	$r = -.058$	PE $\pm$ .064

From the three correlations the inference is quite apparent that the relationship between mechanical ability and intelligence is negligible.

Because the nine year age group was drawn from Classes IIIA to VB inclusive, it was possible to compare the mechanical performance of dull and bright boys of the same chronological age. There were twenty-six boys nine years of age in the bright sections of the Classes IVB, VA and VB and an equal number in the dull sections of the Classes IIIA, IIIB and IVA.

TABLE III.—COMPARISON OF IQ AND MECHANICAL PERFORMANCE OF TWENTY-SIX BRIGHT AND TWENTY-SIX DULL BOYS NINE YEARS OF AGE

	Range of IQ	Median IQ	Range mechanical score	Median mechanical score
Dull boys	73–109	82	38–105	74
Bright boys	100–142	121	45–101	65

Of course twenty-six is too small a group from which to draw conclusions, but this comparison seems to bear out the belief that boys who do not excel in academic subjects would profit by industrial training.

On another occasion the Stenquist Mechanical Assemblying Test Series III was given to a number of the older boys in ungraded classes. From this group sixty boys with mental ages of seven, eight and nine years were chosen. Twenty boys in each mental age group were selected.

The IQ range of these boys was from fifty to seventy. The boys with mental ages of eight and nine years were all thirteen and fourteen years of age chronologically. In the group with mental ages of seven years, a few of the boys were eleven and twelve; the others thirteen and fourteen years of age.

For comparison with the group of ungraded boys a group of sixty normal boys seven, eight and nine years of age chronologically was selected. There were also twenty boys in each age group. Only those with IQ's between ninety and one hundred were chosen.

The scores made by these older boys of inferior intelligence in ungraded classes have been compared with those made by younger normal boys of the same mental age in Table IV.

TABLE IV.—COMPARISON OF THE MECHANICAL SCORES OF SIXTY UNGRADED BOYS AND SIXTY NORMAL BOYS OF THE SAME MENTAL AGE

		Range mechanical score	Average mechanical score
Nine years	Ungraded	10–120	84
	Normal	36–101	64
Eight years	Ungraded	5–120	81
	Normal	35–100	55
Seven years	Ungraded	46–110	78
	Normal	25–72	51

In each group the older boys made a much better showing in their mechanical performance than the younger boys of the same mental age. Could this difference be explained by social experience and interest?

SUMMARY

1. The Stenquist Mechanical Assemblying Test Series III was standardized for boys six to nine years of age.

2. There is a definite relation between the mechanical performance and the chronological age.

3. The correlation between mechanical ability and intelligence is negligible.

4. Dull boys made a higher mechanical score than bright boys of the same chronological age.

5. Older boys of inferior intelligence made a higher mechanical score than younger boys of the same mental age.

A MODIFICATION OF THE DOOLITTLE METHOD: A LOGARITHMIC SOLUTION

R. J. WHERRY

Cumberland University, Lebanon, Tennessee

Much use has been made lately of the Doolittle method of solving normal equations, necessitated by the increased use of the multiple correlation technique. Its more extended use by Psychologists has been urged in a recent article in the JOURNAL OF EDUCATIONAL PSYCHOLOGY.[1]

The purpose of this article is to present a modification of the method making it available, without so much multiplying involved, to those persons who do not have a calculating machine available.

The method as derived by Doolittle is adequately set forth by Mills[2] and needs no elaboration here. Doolittle devised his procedure from the work of Gauss as set forth by Merriman in his text on Least Squares,[3] and it is from this same source that the author has derived the present modification of the Doolittle method.

The one great fault of the Doolittle method, as has been suggested above, is that it requires, for its satisfactory use, the availability of a calculating machine, due to the great amount of multiplying throughout the process. The present paper seeks to obviate this difficulty by substituting the use of logarithms instead, requiring only a five or six place set of logarithms.

In order to illustrate the method let us consider the following example (the one used by Mills to illustrate the Doolittle method is used here for sake of comparison[4]), whose normal equations are:

$$\begin{aligned} 8.3564b_{12.34} + 2.790b_{13.24} + 2.932b_{14.23} + 47.967 &= 0, \\ \pm 6.6645b_{13.24} + 2.063b_{14.23} + 62.039 &= 0, \\ + 7.7893b_{14.23} + 47.519 &= 0. \end{aligned}$$

These equations are of course in the deleted form permitted by the method.

Our problem is to find the values of the terms $b_{12.34}$, $b_{13.24}$, and $b_{14.23}$. All of the necessary work, with the necessary checks is shown in the accompanying table (Table I).

Explanation of Table I.—In the columns A, B, and C appear the coefficients of the "b" values, in column D the constant values, and in column E the sum of all of these previous values, on the basis of which the checking is done.

TABLE I.—SOLUTION OF NORMAL EQUATIONS BY THE WHERRY-DOOLITTLE-GAUSS METHOD

Line	*A* $b_{12.34}$	*B* $b_{13.24}$	*C* $b_{14.23}$	*D* (m)	*E* (s)
I	8.3564	2.790	2.932	47.967	62.0454
II	.....	6.6645	2.063	62.039	73.5565
III	.....	.....	7.7893	47.519	60.3033
1	8.3564	2.790	2.932	47.967	62.0454
2	.92202	.44560	.46716	1.68094	1.79271
3	.00000	9.52358	9.54514	.75892	.87069
4	−1.00000	− .3339	− .3509	− 5.7401	− 7.4248(−7.4249)
5	.....	6.6645	2.0630	62.039	73.5565
6	.....	9.96918	9.99074	1.20452	1.31629
7	.....	− .9315	− .9789	−16.0150	−20.7155
8	.....	5.7330	1.0841	46.0240	52.8410(52.8411)
9	.....	.75838	.03507	1.66299	1.72297
10	.....	.00000	9.27669	.90461	.96459
11	.....	−1.0000	− .1891	− 8.0280	− 9.2170(−9.2171)
12	.....	.....	7.7893	47.519	60.3033
13	.....	.....	.01230	1.22608	1.33785
14	.....	.....	−1.0288	−16.8300	−21.7695
15	.....	.....	9.31176	.93968	.99966
16	.....	.....	− .2050	− 8.7032	− 9.9922
17	.....	.....	6.5555	21.9858	28.5416(28.5413)
18	.....	.....	.81661	1.34215	1.45548
19	.....	.....	.00000	.52554	.63887
20	.....	.....	−1.0000	− 3.3538	− 4.3539(−4.3538)

$$-b_{14.23} + (-3.3538) = 0$$
$$b_{14.23} = -3.3538$$
$$-b_{13.24} + (-.1891)b_{14.23} + (-8.0280) = 0$$
$$b_{13.24} = -7.3938$$
$$-b_{12.34} + (-.3339)b_{13.24} + (-.3509)b_{14.23} + (-5.7401) = 0$$
$$b_{12.34} = -2.0945$$

Check:

$$8.3564b_{12.34} + 2.790b_{13.24} + 2.932b_{14.23} = -47.967$$
$$8.3564(-2.0945) + 2.790(-7.3938) + 2.932(-3.3538) = -47.964$$

In lines I, II, and III are inscribed the coefficients of the normal equations as indicated above. The steps numbered below correspond to the procedure followed in obtaining the results given on the corresponding line in the table.

1. The normal equation I is recopied in line 1.

2. Write the logarithm of each number in line 1 in line 2.

3. Subtract the logarithm found in column A, line 2 from each of logarithms in line 2, recording the result immediately below in line 3. Notice that in the table the log. 9.52358 is written in place of 9.52358 − 10.

4. Look up the anti-logs of the logarithms in line 3 and write in line 4, giving them the sign opposite to the sign of the corresponding number in line 1 when the number in line 1, column A is plus, and the same sign as the numbers in line 1 when the number in line 1, column A is negative. (Check—the sum of the numbers in line 4, columns A to D, must equal the number in column E.)

5. Copy normal equation II on this line.

6. Add the logarithm appearing in column B, line 3 to each of the logarithms in line 2 (columns B to E) and record the results below in line 6.

7. Look up the anti-logs of the logarithms in line 6 and record immediately below in line 7. Notice the sign of the number in column B, line 4 and multiply the signs of the numbers in line 1 by it and record the results before the corresponding numbers in line 7.

8. Add algebraically the numbers in line 7 to those in line 5 and record the results in line 8. (Check—the sum of the numbers in line 8, columns B to D must equal the number in line 8, column E.)

9. Look up and record the logarithm of each number in line 8 immediately below in line 9.

10. Subtract the logarithm in line 9, column B from each of the logarithms in line 9 and record the results in line 10.

11. Look up the anti-logs of the logarithms in line 10 and record in line 11, giving them the sign opposite to the sign of the corresponding number in line 8 if the sign of the number in line 8, column B is plus, and the same sign as the number in line 8 if the sign of the number in line 8, column B is negative. (Check—the sum of the numbers in line 11, columns B to D, must equal the number in line 11, column E.)

12. Copy normal equation III on this line.

13. Add the logarithm appearing in line 3, column C to each of the logarithms in line 2, columns C to E and record the results in line 13.

14. Look up the anti-logs of the logarithms in line 13 and record immediately below in line 14. Notice the sign of the number in line 4, column C and multiply the signs of the numbers in line 1 by it and record the results before the corresponding numbers in line 14.

15. Add the logarithm appearing in line 10, column *C* to each of the logarithms in line 9, columns *C* to *E* and record the results in line 15.

16. Look up the anti-logs of the logarithms in line 15 and record immediately below in line 16. Notice the sign of the number in line 11, column *C* and multiply the signs of the numbers in line 8 by it and record the results before the corresponding numbers in line 16.

17. Add algebraically the numbers in lines 12, 14, and 16 and record the result for each column in line 17. (Check—the sum of the numbers in line 17, columns *C* and *D* must equal the number in line 17, column E.)

18. Look up and record the logarithm of each number in line 17 immediately below in line 18.

19. Subtract the logarithm in line 18, column *C* from each of the logarithms in line 18 and record the results in line 19.

20. Look up the anti-logs of each of the logarithms in line 19 and record in line 20, giving them the sign opposite to the sign of the corresponding number in line 18 if the sign of the number in line 18, column *C* is plus, and the same sign as the corresponding number in line 18 if the number in line 18, column *C* is negative. (Check—the sum of the numbers in line 20, columns *C* and *D* must be equal to the number in line 20, column *E*.)

If there were additional unknowns, the process would be extended in exactly the same manner.

The checks on the various operations have been indicated in the table and in the explanation. The testing of the results immediately following each step reduces the possibility of error to a minimum.

The back solution for the "*b*" values presents no difficulty, being carried out as indicated in the lower part of Table I.

The final check as to the accuracy of the work comes in the substitution of the discovered "*b*" values in one of the complete normal equations. This is done for equation I in Table I.

Conclusions

1. The adaptation of Gauss's method for the solution of normal equations as worked out by Doolittle, while it is very satisfactory for machine calculation, is excessively burdensome when such a machine is not available.

2. The logarithmic solution presented in this paper is based upon exactly the same principles, and gives results of approximately the

same mathematical accuracy, and yet is suited to hand calculation. The author feels that this method will therefore be a boon to many workers in the smaller universities and colleges, whose budgets do not permit the purchase of expensive calculating machines.

References

1. Griffin, H. D.: On Partial Correlation *vs.* Partial Regression for Obtaining the Multiple Regression Equations, *Journal of Educational Psychology*, Vol. XXII, pp. 35–44.

2. Mills, F. C.: "Statistical Methods." Henry Holt & Co., 1924, pp. 577–581.

3. Merriman, M.: "Textbook on Method of Least Squares." John Wiley and Sons, New York. Many Editions.

4. Mills, F. C.: *Op. cit.*, pp. 491–495, 577–581.

NOMOGRAM FOR BLAKEMAN'S TEST FOR LINEARITY OF REGRESSION

HAROLD D. GRIFFIN

Nebraska State Teachers College, Wayne, Nebr.

Blakeman has devised formulas[1] to be applied in testing a correlation ratio and coefficient of correlation to ascertain linearity or non-linearity of regression. There are many versions of Blakeman's tests.[2,3,4,5] The one used in constructing the accompanying nomogram follows Holzinger. For the theory the reader is referred to Blakeman's original article.

The best straight-edge to use in reading this and other nomograms is made by scratching a straight line with a needle on the underside of a strip of good weight celluloid.

---

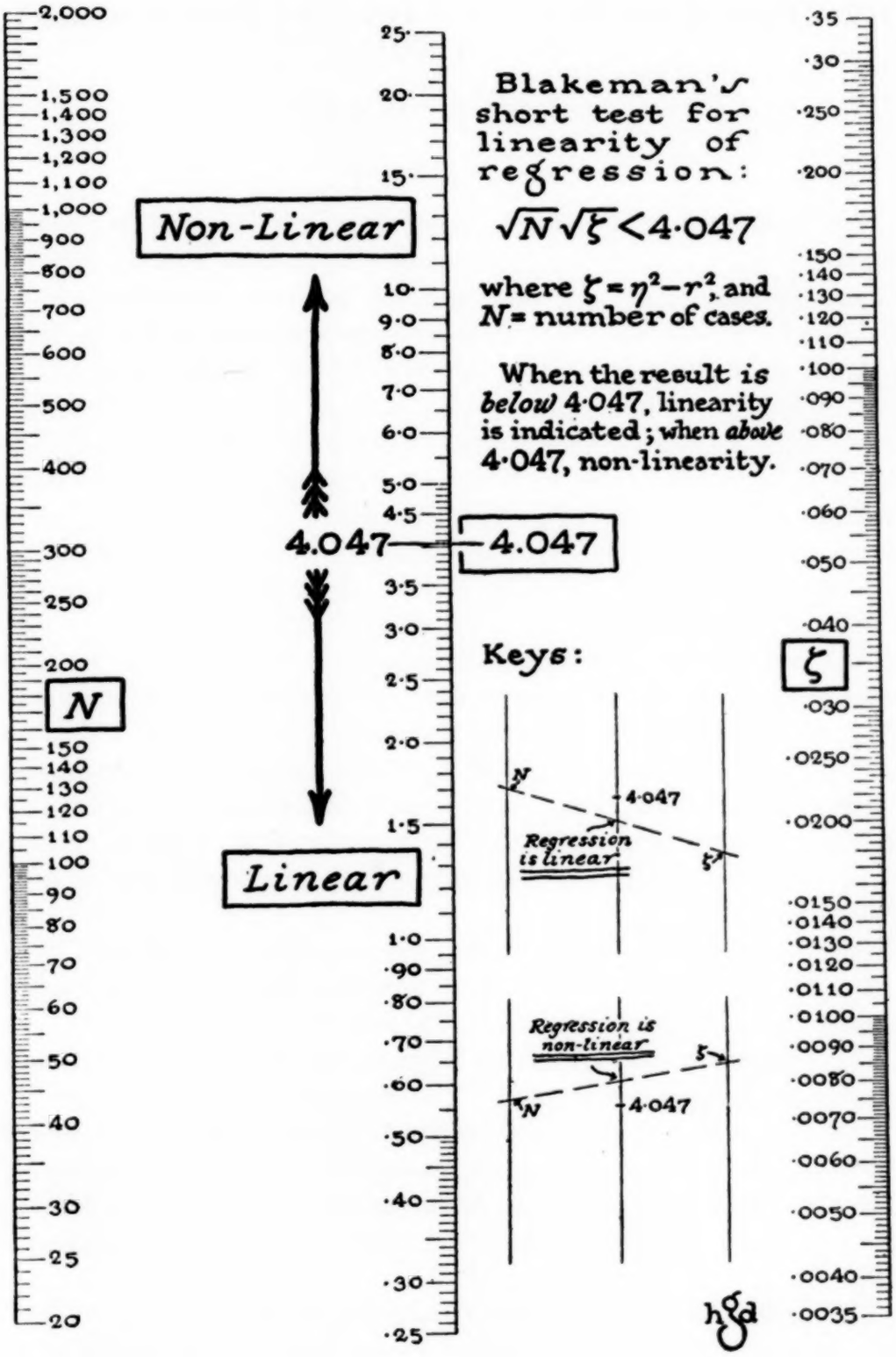
[1] Blakeman, J.: On tests for linearity of regression in frequency distributions. *Biometrika*, Vol. IV, pp. 332–350 (1905).

[2] Blakeman: *Op. cit.*

[3] Garrett, H. E.: "Statistics in Psychology and Education," first edition. Longmans, Green and Co., New York, 1926, pages 209–211.

[4] Gavett, G. I.: "First Course in Statistical Method," first edition. McGraw-Hill Book Co., New York, 1925, p. 239.

[5] Holzinger, K. J.: "Statistical Methods for Students in Education," first edition. Ginn and Co., Boston, 1928, p. 183.


Blakeman's short test for linearity of regression:
$\sqrt{N}\sqrt{\zeta} < 4.047$
where $\zeta = \eta^2 - r^2$, and N = number of cases.
When the result is *below* 4·047, linearity is indicated; when *above* 4·047, non-linearity.
Non-Linear
Linear
4.047
4.047
N
ζ
Keys:
Regression is linear
Regression is non-linear
hgd

THE EFFECT OF PRACTICE ON THE HOMOGENEITY OF A GROUP

ROBERT P. CARROLL

Syracuse University

When the subject of homogeneous grouping is discussed in classes in Education the question sometimes arises as to how long a group will remain homogeneous. As time goes on will the individuals in the group become more and more alike or more and more unlike in ability or skill? Still another question arises and that is, what measure best shows whether the homogeneity of a class or section is decreasing or increasing? In answer to this last question the mean shows nothing. The range has very little significance, for it may remain practically constant while the majority of the members of a group are getting farther apart or closer together in their achievement. The sigma is a very good measure, but it often increases as the average increases. Thus a class may actually be more alike in achievement after a few days of practice on a given function, while the sigma may be much larger than when the practice began. It would seem that the ratio between the mean deviation and the mean, or between the sigma and the mean, that is, the coefficient of variation, would be the best measure of homogeneity. The latter measure was used in connection with two experiments that were conducted in Syracuse University during the last twelve months with groups that were comparatively homogeneous at the outset.

The first experiment consisted of twenty-eight days of drill in one of the broader mental functions,[1] namely, that of silent reading. Changes in the C. V. are shown in Table I.

The changes are by no means regular, nor do they continue consistently in any given direction. The fluctuations are rather similar to those found in general in the results obtained from practice or drill experiments. For the most part, however, the trend is downward, which indicates an increase in homogeneity. This table represents the work of twenty-eight college students in the summer session of 1929.

The second experiment was conducted with the cooperation of thirty-nine students in Educational Psychology in Teachers College.

[1] Drill in Silent Reading for College Students. Reported in *School and Society*, November 9, 1929.

TABLE I.—SHOWING CHANGES IN THE COEFFICIENT OF VARIATION IN SILENT READING. THE NUMBERS HERE REPRESENT HUNDREDTHS BUT ARE GIVEN AS WHOLE NUMBERS TO AVOID THE USE OF DECIMALS

Day of drill	C. V.	Day of drill	C. V.
1	26	15	19
2	23	16	21
3	23	17	18
4	20	18	23
5	12	19	14
6	23	20	15
7	22	21	20
8	22	22	15
9	26	23	14
10	18	24	16
11	19	25	16
12	24	26	12
13	26	27	14
14	17	28	15

TABLE II.—SHOWING THE RESULTS OF THE SECOND EXPERIMENT. THE COEFFICIENTS OF VARIATION ARE AGAIN GIVEN AS WHOLE NUMBERS TO AVOID THE USE OF DECIMALS

Day of drill	C. V.	Day of drill	C. V.
1	24	24	16
2	23	25	16
3	21	26	15
4	17	27	14
5	16	28	14
6	22	29	14
7	19	30	13
8	18	31	15
9	18	32	16
10	18	33	15
11	18	34	16
12	17	35	16
13	17	36	16
14	17	37	13
15	17	38	16
16	18	39	16
17	17	40	15
18	17	41	15
19	16	42	18
20	17	43	15
21	15	44	18
22	15	45	16
23	15		

The purpose of this was to determine the changes in homogeneity as the group was practiced in a rather simple skill. The students were asked to make as many marks as possible in one minute. These were to be made in groups of five, with the left or wrong hand. The drill periods were forty-five in number and extended from October 14 to November 4 in the fall of 1929.

The general trend of the C. V.'s is downward, the changes being very similar to those found in the first experiment. The first C. V. is twenty-four, the last is sixteen, which is only two-thirds as great. After the first twelve trials very few of the coefficients of variation are as much as three-fourths as great as the first two or three. The results of the two studies would seem to indicate that drill in either a broad or narrow mental function would cause the individuals in a comparatively homogeneous group to either remain so, or to become more and more alike in their achievement.

DISCUSSIONS

STATISTICS NEVER EQUIVOCATE

Conrad, in a recent article,[1] attempts to show that the distrust of intelligence ratings arising out of the subjective character of these ratings is unwarranted. He thinks he has done so by showing that the correlation of test scores with ratings is not affected materially by the "correction" of these ratings made by the army psychologists. Obviously Conrad has been misled by the equivocal language used by these psychologist. He quotes as follows:

> . . . If we place all individuals rated "6" for example, in the same class in our contingency table, whether they are all rated by the same person or not, the computed measure of relationship (between ratings and test scores) will be affected by factors other than those relevant to the problem (*viz.*, by the "personal equation") . . . It is necessary, therefore, to equalize the mean ratings of all groups before they are combined. This may be done by treating the class intervals on the rating scale (of each judge) as intervals on the axis of abscissae of a normal probability curve . . .

Now it should be obvious that this so called equalization equalizes the judgment of different judges in range, *and in nothing else.* Thus if one judge tends to go to extremes and rate people either as very stupid or very brilliant whereas another is conservative and tends to place everyone at the golden mean, the army method would "equalize" such differences in temperament. If however one judge considers a certain characteristic, *e.g.*, amenability to discipline, a sign of intelligence whereas another judge can discern intelligence, if present, in minor infractions of regulations, their judgments would not be equalized by the army method. There is in fact no possible method of "equalizing" such differences of opinion. They can only be compromised or averaged. In the special instance, where ratings of the same individuals by different judges are not available, even such averaging is impossible.

With a large number of judges and ratings there is no apriori reason to expect an important change in correlation to arise out of the equalizing adjustment of the army. Nothing whatever follows from the fact that such a change does not occur. The fact that the army psychologists have created the impression that they have corrected for

---

[1] The Bogey of the "Personal Equation" in Ratings of Intelligence. *Journal Educational Psychology*, February, 1932.

the unreliability of personal opinion is merely an illustration of the equivocation so common in the literature. The unreliability of personal opinion is not a bogey, however inconvenient that fact may be to those who try to achieve scientific progress by a determined pull at their boot straps.

CURT ROSENOW.
Institute of Child Guidance, New York City.

"THE EFFECT OF THE 6-22-44-22-6 NORMAL CURVE SYSTEM ON FAILURES AND GRADE VALUES"—A COMMENT

In a recent issue[1] J. DeWitt Davis uses the improved basis for a five-point grading system which I proposed in an earlier issue[2] to show that "a strict application" of it "in a twelve term course would eliminate practically fifty-one per cent of the matriculated group," and that "the values of grades from term to term" would have "little directly comparable meaning." His mathematical and statistical procedure seem to be entirely sound and his conclusions inevitable, once his fundamental assumptions are granted. That these assumptions are valid, however, may properly be questioned.

1. Any division of the area under the normal probability curve may be made, and none necessarily implies any per cent of "failures." The lower six per cent in the system which I suggested simple means "distinctly inferior" in comparison with the remainder of the group. An institution or an individual instructor or department may or may not decide that part or all of this inferior group should be "failed." In many institutions both conditions and failures are included in this group. Yet the conditions and sometimes the failures may be made up. They do not necessarily imply permanent failure and elimination from school.

2. Even more fundamental, however, is Mr. Davis's assumption that failure in a single course automatically results in elimination from school at the end of the term in which it occurs. The average student takes from three to five courses; regulations in institutions vary widely

---

[1] Davis, J. DeWitt: The Effect of the 6-22-44-22-6 Normal Curve System on Failures and Grade Values. *Journal of Educational Psychology*, Vol. XXII, No. 8, November, 1931, pp. 636–640.

[2] Eells, Walter Crosby: An Improvement in the Theoretical Basis of the Five Point Grading System Based on the Normal Probability Curve. *Journal of Educational Psychology*, Vol. XXI, No. 2, February, 1930, pp. 128–136.

as to the number of these which must be failed to result in probation or complete disqualification. At Stanford University, for example, according to the latest published report of the Scholarship Committee, the disqualifications for the autumn quarter, 1929–1930, were fifty-nine in an undergraduate student body of two thousand six hundred ninety-six, or approximately two per cent, although the number of conditions and failures in separate courses was much higher. This assumption that a condition or failure in one subject any quarter automatically disqualifies a student from school or college is the fundamental fallacy that invalidates Mr. Davis's major conclusions.

3. On the assumption that elimination would take place on the foregoing basis, Mr. Davis constructs a nomograph and a table to show that sophomore grades would not be at all equivalent to freshman ones, and junior and senior grades decreasingly so. Grades received at these different levels, however, are not measures of a constant initial ability, but are measures of achievement at the different levels. Assume a normal distribution of grades in a freshman class in trigonometry. Probably a sophomore class in calculus would all be above the median of the trigonometry class, if marked on the basis of a comparable examination in trigonometry; but the sophomore grading should be adjusted to the level of sophomore achievement in the particular subject. An "A" grade in calculus would mean that a student was among the upper six per cent of the average students taking calculus. He should not be judged by a freshman trigonometry standard. The senior should be able to secure superior grades in most if not all freshman subjects. He is not therefore entitled to "A" grades in every subject. Inferiority or superiority in college grading can hardly mean with reference to a fixed standard; it must be relative. One is superior or inferior with reference to his peers in the same field.

4. Incidentally Mr. Davis is in error when he says with reference to the suggested 6-22-44-22-6 system that practically one hundred per cent is included within ± 3 sigma range. The absolute error is not serious, but the principle is important. One of the distinctive theoretical as well as practical advantages of the system which I proposed was that it makes no such arbitrary assumption of the length of the base line of the normal probability curve. An assumption of a range of ± 3.0 sigma leads to a standard of 3.45 per cent of "A" grades, while a slight change in this assumption to ± 2.5 sigma leads to a standard of 6.06 per cent instead—a striking difference for such a slightly differ-

ent assumption.[1] One of the great virtues of the suggested 6-22-44-22-6 system is that it makes no such arbitrary assumption but includes the entire range of the normal probability curve as well as assuring the equality of the distances between the means of the successive grade groups.

WALTER CROSBY EELLS.

School of Education, Stanford University, California.

[1] Rugg, Harold O.: "Statistical Methods Applied to Education." Houghton Mifflin Co., Boston, 1917, pp. 217–218.

BOOK REVIEWS

NORMA V. SCHEIDERMANN. *Psychology of Exceptional Children.* Pp. XVII + 520. New York: Houghton Mifflin Co., 1931.

As Cubberly says in his introduction to *Psychology of Exceptional Children,* "there are thousands of children in our schools today who are realizing but a small percentage of their educational possibilities." Miss Scheidermann's book is an effort to present in a single volume pertinent available material concerning the atypical child, whether he be afflicted with deafness, blindness, feeble-mindedness, or what-not. It is a valiant attempt, but predestined to superficiality by the very great scope of her subject.

Presumably the author did not have space for a careful critical evaluation of many of the investigations to which she refers; otherwise, it is difficult to understand how she could offer, without adverse comment, such generalizations as those of Murchison. (See quotation on page 458.) In discussing the relationship between church membership and criminality, Murchison has made such dogmatic statements as "Methodists and Hebrews are superior as plain thieves," offering as proof a series of unreliable percentages based, in some instances, upon fewer than a dozen cases. Data and quotations from such questionable sources appear altogether too often in Miss Scheidermann's book, while there is a frequent failure to present the conclusions arrived at by men who are outstanding in their field. For example, Garth is not even referred to in the section on intelligence and race . . . And this reviewer is still puzzling over the possible justification for giving more space to a discussion of the left-handed child than to either the mentally subnormal or the mentally gifted.

In spite of its many weaknesses, it is felt that *Psychology of Exceptional Children* should be given a place as a reference book, since Miss Scheidermann has here made easily accessible a number of facts concerning the child who deviates in any particular from the normal.

HERBERT A. CARROLL.

University of Minnesota.

JAMES DUNLAP LICKLEY. *The Nervous System.* Pp. XII + 144. New York: Longmans, Green and Company, 1931.

In the new edition of *The Nervous System* Dr. Lickley has incorporated some new material but in scope, style and organization there is no

change from the previous edition. It is a clear and concise introductory handbook of the anatomy and physiology of the nervous system intended for non-medical students. The dominant level of description is morphological; physiological facts are included merely to clarify the details of structure. There are one hundred seventeen well selected and well presented illustrations. Most of these are from 17th and the 24th editions of *Gray's Anatomy*. There are ten chapters in the book. In the first two the nature of nerve cells and nerve fibres are considered. In the last two localization areas and the sense organs, respectively, are treated. The remaining chapters are devoted to a description of the structure and function of the various parts of the nervous system. The work of Lashley, Child and Herrick are not considered. Here and there the author adds a sentence or two suggesting newer findings. The very simplicity and conciseness make extensive approaches impossible. Hence in his discussions the author wisely adopts the method indicated in his treatment of theories of color vision. Without mentioning other theories he prefaces a brief discussion of Hering's theory, with a statement which says in effect that the one he selected seems to cover many of the known facts. The book contains an index.

H. MELTZER.
Psychiatric Child Guidance Clinic, St. Louis.

DOUGLAS A. THOM. *Normal Youth and Its Everyday Problems.* Pp. 367. New York: D. Appleton & Co., 1932.

In *Everyday Problems of the Everyday Child* Dr. Thom achieved recognition for his excellent treatment of the problem of normal children in the first decade of life. His second book considers the behavior problems of normal adolescents and keeps up the excellent standard which he set in the first. He combines with his broad experience a gift for clear and interesting exposition.

For parents, teachers, and others interested in youth, Dr. Thom says he has attempted "to discuss some of the pitfalls and hazards that surround the adolescent and in so far as possible, to suggest how they may be avoided; or, in those cases where difficulties already exist, to indicate in a general way the approach most likely to be helpful in the solution of the problem." And the everydayness of the problems should be stressed here; the normal, not abnormal, youth is the center of the picture. There are those who will find in this book the first explanations of some of their own adolescent difficulties, but whether

that is true or not, it is my feeling that the book cannot be read without increased tolerance and concern for youths of high school and college age.

For the beginner, abstract discussions of behavior problems are difficult to apply to concrete cases. Dr. Thom has avoided this objection through the inclusion of numerous definite illustrations of the type of problem discussed. They are succinct, omitting extraneous details, and lend a reality to the discussion which most readers will appreciate.

If a reviewer may be permitted a bias, it leans to the chapters on Adjustment to Maturing Sex Drives, and Education. Physically the book is attractive, well-bound and contains an index.

J. H. COLEMAN.

Huntington, N. Y.

L. T. TROLAND. *The Principles of Psychophysiology*, Vol. III. Pp. XVII + 446. New York: D. Nostrand and Co., 1932.

CARL MURCHISON, Editor. *A History of Psychology in Autobiography*, Vol. II. Pp. 407. Worcester: Clark University Press, 1932.

D. H. KULP, II. *Educational Sociology*. Pp. XII + 591. New York: Longmans, Green and Co., 1932.

A. J. NOCK. *The Theory of Education in the United States*. Pp. 160. New York: Harcourt, Brace, 1932.

Here we have four most significant and interesting books on psychology and education. The third volume of Dr. Troland's monumental work on psychophysiology is bound to be welcomed by all psychologists with gratitude, as have the two previous volumes. All one can do is express gratitude to the author for a difficult job magnificiently done.

The second volume of the Clark University Press series of professional autobiographies of continental and American psychologists of note makes as fascinating reading as did its predecessor. Students of the sciences need just this personal touch, this acquaintance with the human beings behind science, their struggles, failures and triumphs in the disinterested search for truth, to make them realize the human significance of science and to whet their appetite for making themselves familiar with its accomplishments. The Clark University

Press is doing psychology, and science in general, a great service in sponsoring such an enterprise.

Dr. Kulp overworks the term *scientific* and makes the reader weary with ninety-nine pages of repetitions that education has not been, should be, and could be scientific. A brief introductory chapter would have been quite sufficient to make his point. There is hardly any need or justification for belaboring those educators of the past who theorized on the nature and function of education when they had no scientific data on which to base a science of education. And from the way Dr. Kulp uses the term *philosophy* to embrace everything towards which he is prejudiced, one might suspect that his conception of philosophy is hardly scientific. But once Dr. Kulp has worked his spite out of his system he makes a most significant contribution to the sociological aspect of education. His book should attract the closest attention of students of both education and sociology.

To the reviewer, A. J. Nock's book of one hundred sixty pages gives the most suggestive, succinct and illuminating survey and analysis of the theoretical basis of education in the United States now in print. Mr. Nock offers no panaceas for our educational ills, for which one is grateful to him. But he does point out that the three fundamental ideas in our educational theory, namely, equality, democracy, and the bearing of literacy upon public order and honest government, stand in need of careful clarification and modification. No one interested in education can afford to miss this most readable and honest book.

MAX SCHOEN.

Carnegie Institute of Technology.

MARION MONROE. *Children Who Cannot Read.* Pp. XVI + 205. Chicago: The University of Chicago Press, 1932.

Evidence accumulated during the past ten years in the field of reading disabilities indicates that schools can no longer neglect their responsibility to the non-reader. With techniques now available all cases of reading deficiency may be partially or fully remedied providing that they are not entirely due either to lack of mental ability or to constitutional emotional deviations.

The present monograph, an outgrowth of the previous interests of the author, is an important addition to the literature on diagnostic and remedial reading. The report consists essentially of methods of quantitative and qualitative measurement of reading ability, a discussion of causative factors, and results of remedial treatment. Four

hundred fifteen children with special reading defects were compared with a control group of one hundred one. It is unfortunate that the control group used was of superior intelligence in comparison to the normal or sub-normal experimental group since, in general, reading ability is markedly influenced by mental ability. Monroe's earlier method of diagnosis has been amplified somewhat in the present study. A reading index, derived by dividing the reading grade (achievement) by an average grade placement involving (1) chronological age, (2) mental age, and (3) arithmetical ability. It is difficult to understand why arithmetical accomplishment is included in this reading index. In fact, as shown by the two illustrations cited on p. 14, an average of grade placements derived from mental and chronological ages is almost identical with the average when arithmetic is included. The unnecessary use of the arithmetical score is born out by the author's own conclusions in which no mention is made of arithmetic in listing functions upon which progress under remedial instruction depends.

The chapter on "causative factors in reading defects" is the least adequate part of a book which is highly commendable on the whole. In this section the interpretations of the author have been influenced perhaps by certain points of view which need to be modified in the light of inferences which may be derived from a comprehensive analysis of contributions to the psychology of reading. An example is the assumption that left-eyedness fosters (1) excessive regressions in reading and (2) reversals in perception of letters and single words. One is more justified in inferring that both left-eyedness and tendencies to sinistrad sequencies in word and letter perception are due to cerebral dominance. Eye movements tend to be guided by perceptual tendencies rather than vice versa.

"The rate of progress in reading under remedial instruction was found to be a function of the child's intelligence, his age, the number of hours spent in training, the number of months during which treatment was continued, the severity of the disability, the personality and behavior difficulties encountered in applying the remedial training, and the closeness of supervision of the remedial techniques."

This book will be welcomed by educators and clinical psychologists. Diagnosis and remedial treatment are clearly stated and have proved practical and effective for ninety-eight per cent of all cases tutored under close supervision made normal or accelerated progress in reading.

MILES A. TINKER.

University of Minnesota.

DE LA FONTAINERIE. *French Liberalism and Education in the Eighteenth Century.* Pp. VII + 385. New York: McGraw-Hill Book Company, 1932.

This book contains the translations of the following essays on education in France in the eighteenth century: La Chalotais' *Essay on National Education,* Turgot's *Memorial on Education,* Diderot's *Plan of a University for the Russian Government,* and Condorcet's *Report on the General Organization of Public Instruction.* The authors of these essays register a complaint against the deplorable condition of education in France of that period. Chalotais centers his attack on the clergy. He wants education to be the function of the State. Turgot pleads for equality of educational opportunity. Diderot's polemic is almost a satire on educational conditions in France, although it was written as a plan of a university for Russia. He decries the undue emphasis on the study of the Latin language and classical literature at the expense of the practical sciences. Condorcet, like other liberal reformers of his day, does not conceal his animosity toward the church and the classics, although his essay indicates that he proposed a milder means of reforming education than his contemporaries. He urges the State to assume the responsibility of training its citizens in civic duties and in moral obligations to society.

If one were not familiar with the circumstances which so profoundly influenced Rousseau's writing of *Emile,* he might have occasion to believe that the writers of the essays on education compiled by Mr. De La Fontainerie were inspired through some personal or selfish motive. But such was not the case. They had witnessed and deplored the abuses and misuses of educational practice by the churchmen of their time, and were thus inspired to suggest a remedy for the educational ills of the period. Their sole aim was to free the masses of people from the bondage to eccliasticism and to offer possible measures for a State-controlled system of education. Aside from the prejudices and oddities with which the essays are, in part, characterized, the discourse contains some very constructive criticism of educational practice, even in modern times.

The reviewer believes that the usefulness of the book might have been improved by introducing, at certain points in the translations, a few contrasts and comparisons with corresponding educational practices in our own schools. A few parallelisms might have been indicated for the sake of students in courses of history of education without

detracting greatly from the organization of the book, and possibly without disrupting the discourse excessively.

The translator has made a valuable contribution to the entire field of education, for there is much in the modern theory of education anticipated in the work. Students of education, teachers, supervisors, and administrators will find this compilation of essays very inspirational and helpful in their study and service. The book is translated in simple and effective English with an introduction extensive enough to furnish the ordinary reader with a good background of the historical facts incident to the study and understanding of the essays.

ROBERT G. SIMPSON.

Carnegie Institute of Technology.

G. A. YOAKAM. *The Improvement of the Assignment.* Pp. VII + 398. New York: Macmillan Company, 1932.

One is impressed at his first glance into this book by the amount of useful information which the author has supplied on an apparently limited subject. Yet it would not be an exaggeration to say that it would have been possible to have included considerably more material of a very important character in the book. The subject is a worthy one for such a thorough treatment as the author has given it.

The Improvement of the Assignment covers an extensive range of activities involved in the teaching and learning processes. Some of the more important ones which are considered in the book are: the importance of the assignment, the need of improved assignments, the characteristics of a good assignment, the conditions essential for making successful assignments, the motivation of the assignment, the preparation of new-type assignment materials, and the evaluation of the assignment. These important phases of the assignment are discussed in connection with the experimental evidence which the author has been able to obtain from time to time in his classes and seminars, and on the basis of evidence collected from studies and investigations which have been reported in research bulletins, educational periodicals, and in other sources.

The author furnishes much information which clears up many of the controversal points of making assignments. He is careful in drawing conclusions from the research data which he has occasionally used and in suggesting reforms which cannot be, in part, supported by good practice. His discussions are not merely based on personal opinion. They are usually substantiated by such objective data as are available

in educational literature. He is conservative in expressing his opinion on matters which are, as yet, controversal. One of the most important contributions which Dr. Yoakam makes is this objective point of view in his book.

A special feature of the book is the Chapter dealing with practical suggestions on the preparation of new-type assignment materials. A technique, together with illustrations, models, and useful suggestions for the guidance of the beginner in the writing of assignments, is indicated. This is the first treatment of the preparation of new-type materials in any book with which the reviewer is familiar.

The book is explicitly written in good English and is very easy to read. It should be found an invaluable source of information for those who are engaged in planning and organizing activities and in directing study.

ROBERT G. SIMPSON.
Carnegie Institute of Technology.

SAMUEL W. HARTWELL. *Fifty-five "Bad" Boys.* Pp. XVIII + 359. New York: Alfred A. Knopf, 1931.

In this book the present director of the Worcester Child Guidance Clinic describes and interprets his psychotherapeutic endeavors with fifty-five difficult behavior problem boys. The cases reported represent a chronological series studied and treated by the author at the Judge Baker Foundation during "a certain period of five months." The introduction, by Dr. William Healy, is, therefore, in order. The content is presented under three divisions. Part One consists of three short chapters describing the problem, the plan, and technique respectively. Almost three-fourths of the pages are devoted to a description of the boys. This is done in Part Two. The case studies are frequently presented under such appealing captions as "Pleasures from Delinquency," "Distorted Lives," "Misunderstanding." The three very short chapters in Part Three include a rough statistical summary and evaluation of the cases and outcomes.

Individuals in search of profound analysis of deeper meanings underlying behavior, as orthodox psychoanalysts talk of such things, will find practically none of it in this book. Neither will they find obscurantism demonstrated. The case studies are not purported to be models for social workers to follow. Characteristically the boys are described in a manner which beginning students in a course in mental hygiene have described as concise, clear and intelligible. Some well

trained psychiatrists who expect more from published case studies than stereotyped presentations which follow a pre-determined pattern of thinking about, thinking with and thinking for every case find the descriptions, tiresome and interpretations, illogical. In reading the book one does, at times, hit spots which seem to justify Dr. Healy's characterization of the author as possessing a "rugged honesty" as well as the qualities necessary for a sympathetic understanding of human beings. But, more often than not, it is the kind of sympathetic understanding and ruggedness which the reviewer associates with the experiential background of "educational" camp directors rather than with psychiatric or psychological learning and training. That is, with advisors who talk of the need for giving oneself rather than understanding oneself and whose concern is with "making" or "breaking" a boy rather than understanding him and maintaining a learning attitude which will make further understanding possible.

A feature which will interest interviewers is the discussion of degrees of rapport to adopt in working with different boys. The range described is from "friendly belief" to "dependent attachment." The two intermediate degrees are labelled personal trust and "personality contact." In his treatment of this topic the author fails adequately to consider mood or temperamental differences in the same individual in different sessions and sometimes in the same session.

H. MELTZER.

Psychiatric Child Guidance Clinic, St. Louis.

PUBLICATIONS RECEIVED

RAYMOND WALTER BIXLER: *Anglo-German Imperialism in South Africa*, 1880–1900. Baltimore: Warwick & York, Inc., 1932, pp. 181.

J. HERBERT BLACKHURST: *Introducing Education*. New York: Longmans, Green & Co., 1932.

GEORGE JOSEPH CAIRNS: *An Analytical Study of Mathematical Abilities*. Washington: Catholic University of America, 1932, pp. 105 (paper).

LUCILE CHASE: *Motivation of Young Children*. Iowa City: University of Iowa, 1932, pp. 119 (paper).

ALBERT BEECHER CRAWFORD and STUART HOLMES CLEMENT, Editors. *The Choice of an Occupation*. New Haven: Yale University, 1932, pp. 495 (paper).

EDGAR A. DOLL, Editor. *Twenty-five Years: A Memorial Volume in Commemoration of the Twenty-fifth Anniversary of the Vineland Laboratory*. Vineland, N. J.: Training School at Vineland, 1932, pp. 135.

F. M. EARLE: *Methods of Choosing a Career*. London: George C. Harrap & Co., Ltd., 1931, pp. 334.

HENRY P. EMERSON and IDA C. BENDER: *Modern English*. Book One, pp. 249. Book Two, pp. 342. Book Three, pp. 406. Revised by Blanche Jennings Thompson. New York: The Macmillan Co., 1932.

ROBERT H. GAULT. *Criminology*. New York: D. C. Heath & Co., 1932, pp. 461.

HUBERT GRUEBDER: *Experimental Psychology*. Milwaukee: Bruce Pub. Co., 1932, pp. 455.

CLYDE HISSONG: *The Activity Movement*. Pp. x, 122. Baltimore: Warwick and York, Inc., 1932.

WENDELL JOHNSON: *The Influence of Stuttering on the Personality*. Iowa City: University of Iowa, 1932, pp. 140 (paper).

MARTIN KEILHACKER: *Der Ideale Lehrer Nach der Auffassung der Schüler*. Freiburg, Germany: Herder & Co., 1932, pp. 156 (paper).

DAVID KENNEDY-FRASER: *Education of the Backward Child*. New York: D. Appleton & Co., 1932, pp. 236.

EDWIN A. KIRKPATRICK: *The Sciences of Man in the Making*. New York: Harcourt, Brace & Co., 1932, pp. 396.

LEONARD V. KOOS and GRAYSON N. KEFAUVER: *Guidance in Secondary Schools*. New York: The Macmillan Co., 1932, pp. 640.

DANIEL H. KULP, II: *Educational Sociology*. New York: Longmans, Green & Co., 1932, pp. 590.

J. H. LANDMAN: *Human Sterilization*. New York: The Macmillan Co., 1932, pp. 341.

CHARLES W. E. LEIGH: *Catalogue of the Library for Deaf Education*. Manchester, England: The Manchester University Press, 1932, pp. 143.

ALBERT MANSBRIDGE: *Margaret McMillan: Prophet and Pioneer*. London: J. M. Dent & Sons, Ltd., 1932, pp. 178.

CHARLES HAROLD MCCLOY: *The Measurement of Athletic Power*. New York: A. S. Barnes & Co., 1932, pp. 178.

EDWARD J. v. K. MENGE: *Jobs for the College Graduate in Science*. Milwaukee: The Bruce Publishing Co., 1932, pp. 175.

A. K. MILLER: *The Mixochoanitic Cephalopods.* Iowa City: University of Iowa, 1932, pp. 86 (paper).

MARION MONROE: *Children Who Cannot Read.* Chicago: University of Chicago Press, 1932, pp. 205.

WALTER S. MONROE and RUTH STREITZ: *Directing Learning in the Elementary School.* Garden City, N. Y.: Doubleday, Doran & Co., Inc., 1932, pp. 480.

CHARLES W. MORRIS: *Six Theories of Mind.* Chicago: University of Chicago Press, 1932, pp. 337.

JAY B. NASH: *Character Education through Physical Education.* New York: A. S. Barnes & Co., 1932, pp. 315.

LAURANCE A. PETRAN: *An Experimental Study of Pitch Recognition.* Psychological Monographs, Volume XLII, No. 6. Princeton: Psychological Review Co., 1932, pp. 124 (paper).

EVELYN BUTLER PHILLIPS: *Curricula of the Small High Schools of Maine.* Orono, Maine: University Press, 1932, pp. 89.

FRANCIS F. POWERS: *Character Training.* New York: A. S. Barnes & Co., 1932, pp. 158.

BENJAMIN RAND: *Berkeley's American Sojourn.* Cambridge: Harvard University Press, 1932, pp. 79.

G. W. REAGAN: *Fundamentals of Teaching.* Chicago: Scott, Foresman & Co., 1932, pp. 554.

WALTER C. RECKLESS and MAPHEUS SMITH: *Juvenile Delinquency.* New York: McGraw-Hill Book Co., Inc., 1932, pp. 412.

CHRISTIAN A. RUCKMICK, Editor: *University of Iowa Studies in Psychology,* No. XV. Princeton, N. J.: Psychological Review Co., 1932, pp. 303 (paper).

ROBERT R. RUSK: *Research in Education: an Introduction.* London: University of London Press, 1932, pp. 108.

LOUISE C. SEIBERT: *A Series of Experiments on the Learning of French Vocabulary.* Baltimore: Johns Hopkins Press, 1932, pp. 106 (paper).

CHESTER E. SIPPLE: *British Foreign Policy Since the World War.* Iowa City: University of Iowa, 1932, pp. 157 (paper).

CALVIN P. STONE, CHESTER W. DARROW, CARNEY LANDIS, and LENA L. HEATH: *Studies in the Dynamics of Behavior.* Chicago: University of Chicago Press, 1932, pp. 332.

The Direct Contribution of Educational Psychology to Teacher Training. Yearbook Number XX of The National Society of College Teachers of Education. Chicago: University of Chicago Press, 1932, pp. 154 (paper).

The Vibrato: Studies in the Psychology of Music, Volume I. Iowa City, Iowa: University of Iowa, 1932, pp. 382 (paper).

DOUGLAS A. THOM: *Normal Youth and Its Everyday Problems.* New York: D. Appleton & Co., 1932, pp. 368.

EDWARD L. THORNDIKE and OTHERS: *The Fundamentals of Learning.* New York: Teachers College, Columbia University, 1932, pp. 638.

LEONARD T. TROLAND: *The Principles of Psychophysiology.* New York: D. Van Nostrand Co., 1932, pp. 446.

JAMES G. UMSTATTD: *Student Self-support at the University of Minnesota.* Minneapolis: University of Minnesota Press, 1932, pp. 205.

DOROTHY VAN ALSTYNE: *Play Behavior and Choice of Play Materials of Preschool Children.* Chicago: University of Chicago Press, 1932, pp. 104 (paper).

RAYMOND HOLDER WHEELER: *The Laws of Human Nature: A General View of Gestalt Psychology.* New York: D. Appleton & Co., 1932, pp. 235.

ROTHWELL WILCOX: *Private Secondary Education in the Association of Colleges and Secondary Schools of the Southern States.* Baltimore: Johns Hopkins Press, 1932, pp. 151 (paper).

WILLIAM L. WRINKLE and WINFIELD D. ARMENTROUT: *Directed Observation and Teaching in Secondary Schools.* New York: The Macmillan Co., 1932, pp. 399.